Optimal Capital Taxation in an Economy with Innovation-Driven Growth

Chen, Ping-ho and Chu, Angus C. and Chu, Hsun and Lai, Ching-Chong

Tunghai University, University of Macau, Tunghai University, Academia Sinica

February 2019

Optimal Capital Taxation in an Economy with Innovation-Driven Growth*

Ping-ho Chen

Department of Economics, Tunghai University, Taiwan

Angus C. Chu

University of Macau, Macau, China

Hsun Chu

Department of Economics, Tunghai University, Taiwan

Ching-chong Lai

Institute of Economics, Academia Sinica, Taiwan

Department of Economics, National Cheng Chi University, Taiwan

Institute of Economics, National Sun Yat-Sen University, Taiwan

Department of Economics, Feng Chia University, Taiwan

August 2021

***Acknowledgements**: We are deeply grateful to the Editor, William Barnett, an anonymous associate editor, and two anonymous referees for their insightful comments and suggestions. The paper has also benefited from useful comments by Been-lon Chen, Fu-sheng Hung, and Ming-fu Shaw. Angus Chu gratefully acknowledges financial support from the Asia Pacific Academy of Economics and Management at the University of Macau. The usual disclaimer applies.

Optimal Capital Taxation in an Economy with Innovation-Driven Growth

Running head: Optimal capital taxation and R&D growth

Correspondence: Hsun Chu, Department of Economics, Tunghai University, No.1727, Sec.4, Xitun Dist., Taiwan Boulevard, Taichung, 40704 Taiwan. Tel.: +886-4-23590121 #36119; Email: hchu0824@gmail.com

Optimal Capital Taxation in an Economy with Innovation-Driven Growth

Ping-ho Chen
Department of Economics, Tunghai University, Taiwan
Angus C. Chu
University of Macau, Macau, China
Hsun Chu
Department of Economics, Tunghai University, Taiwan
Ching-chong Lai
Institute of Economics, Academia Sinica, Taiwan
Department of Economics, National Cheng Chi University, Taiwan
Institute of Economics, National Sun Yat-Sen University, Taiwan
Department of Economics, Feng Chia University, Taiwan

August 2021

Abstract

This paper investigates optimal capital taxation in an innovation-driven growth model. We examine how the optimal capital tax rate varies with externalities associated with R&D and innovation. Our results show that the optimal capital tax rate is higher when (i) the "stepping on toes effect" is smaller, (ii) the "standing on shoulders effect" is stronger, or (iii) the extent of creative destruction is smaller. The optimal capital tax rate is more likely to be positive when there is underinvestment in R&D. Moreover, the optimal capital tax rate and the monopolistic markup exhibit an inverted-U relationship. By calibrating our model to the US economy, we find that the optimal capital tax rate is positive, at a rate of around 6.6%. Finally, we consider a number of extensions and find that the result of a positive optimal capital tax is robust.

Keywords: Optimal capital taxation, R&D externalities, innovation

JEL classification: E62, H21, O31

1 Introduction

Capital income is taxed worldwide. The estimated effective average tax rates on capital income are around 40% in the United States and 30% in EU countries. In some countries, such as the United Kingdom and Japan, the capital income tax rates are even up to nearly 60%. From the perspective of welfare maximization, whether these capital tax rates are too high or too low is an important policy question.

Despite the fact that capital taxes are commonly levied in the real world, a striking theory put forth by Judd (1985) and Chamley (1986) suggests that the government should only tax labor income and leave capital income untaxed in the long run. A number of subsequent studies, including Chari et al. (1994), Jones et al. (1997), Atkeson et al. (1999), and Chari and Kehoe (1999), relax key assumptions in Judd (1985) and Chamley (1986), and find their result to be quite robust. The idea of a zero optimal capital tax has then been dubbed the Chamley-Judd result, which turns out to be one of the most well-established and important results in the optimal taxation literature.[1]

In this paper, we revisit the Chamley-Judd result in an innovation-driven growth model. There are several reasons as to why we choose this environment to study optimal taxation. First, as stressed by Aghion et al. (2013), it appears that the consideration of growth does not play much of a role in the debate on the Chamley-Judd result. However, given that the recent empirical evidence suggests that the tax structure has a significant impact on economic growth (e.g., Arnold et al., 2011), it is more plausible to bring the role of growth into the picture. Second, along the line of the optimal taxation literature, production technology is treated as exogenously given. The role of endogenous technological change driven by R&D has thus been neglected in previous models. In view of the fact that innovation is a crucial factor in economic development as well as in the improvement of human well-

[1] More recently, Chari et al. (2020) further support the result that capital should not be taxed by extending the model to include richer tax instruments that the government can access.

being, overlooking this element could lead to a suboptimal design of tax policies. Our study thus aims to fill this gap. Third, as pointed out by Domeij (2005), a key premise in early contributions supporting the Chamley-Judd result is that there exist no inherent distortions and externalities in the economy. If market failures are present, the optimal capital income tax might be different from zero. Thus, we introduce an innovation market that featues various R&D externalities put forth by Jones and Williams (2000). Within this framework, we can study how the optimal capital taxation and R&D externalities interact in ways not so far understood.

By calibrating the model to the US economy, our numerical analysis shows that the optimal capital income tax rate is around 6.6%. The reason for a positive optimal capital income tax in our R&D-based growth model can be briefly explained as follows. In essence, the Chamley-Judd result involves a tax shift between capital income tax and labor income tax. The basic rationale behind a zero optimal capital tax is that taxing capital generates more distortion than taxing labor, because taxing capital creates a dynamic inefficiency for capital accumulation. In our R&D-based growth model, by contrast, innovation requires R&D labor, as typically specified in standard R&D-based growth models (e.g., Romer, 1990; Jones, 1995; Acemoglu, 1998).[2] Under such a framework, taxing labor has a detrimental effect on the incentives for innovation and growth. This introduces a justification for taxing capital income instead of labor income. On these grounds, it might be optimal to have a non-zero capital income tax rate.

Although the result of a positive capital income tax rate is not new in the literature, our study provides insights by examining with what features of the innovation process would the

[2] There are two specifications regarding the innovation process in typical R&D-based growth models: the knowledge-driven specification (i.e., R&D using labor/scientists as inputs) and the lab-equipment specification (i.e., R&D using final goods as inputs). Our analysis adopts the former approach by following the viewpoint of Romer (1990) and Jones (1995) and also the empirical viewpoint of Einiö (2014) who points out that R&D is a labor-intensive activity. If we instead adopt the lab-equipment specification, the numerical values of the optimal capital tax rate would be different. However, the nature of the relationships between R&D externalities and the optimal capital tax, which is our central goal in this paper, will not change.

optimal capital tax rate be positive. By varying the parameters capturing important R&D externalities to see how the optimal capital income tax responds, our analysis reveals the following findings. First, under the benchmark parameters, the optimal capital tax rate is positive, but this result can be sensitive to the parameter that determines the monopolistic markup. Second, when knowledge spillovers are large or R&D duplication externalities are small (thereby increasing the chances of underinvestment in R&D), it is more likely that a positive optimal capital income tax rate will result. Third, when creative destruction is more relevant in the R&D process, the optimal capital income tax rate should be lower. Fourth, a higher government spending ratio pushes toward a positive optimal capital income tax.

Another contribution of this paper is that we identify the role of the monopolistic markup played in determining optimal capital taxation. Our numerical analysis shows that the optimal capital income tax and the markup display an inverted-U shaped relationship. In existing studies, a well-known result is that when the intermediate firms are imperfectly competitive, capital investment is too low compared to the socially optimal level (e.g., Aiyagari, 1995; Judd, 1997, 2002; Coto-Martínez et al., 2007). Accordingly, the government should subsidize capital income to induce a higher level of capital investment, implying that the optimal capital income tax tends to decrease when the monopolistic markup increases. In addition to capturing this traditional effect, our present R&D-based growth model also discloses another effect. In our model, the markup is inversely determined by the elasticity of substitution between intermediate goods. A reduction in the substitution elasticity that raises the markup amplifies the productivity of differentiated varieties in the production of final goods and hence increases the social value of R&D. As a result of this, the government is inclined to subsidize labor by taxing capital given that the R&D sector uses labor. In consideration of this R&D effect, an increase in the monopolistic markup is not necessarily accompanied by a lower optimal capital income tax.

Related Literature. There is a vast literature that attempts to overturn the Chamley-Judd result. Most of these papers obtain positive long-run capital taxation by changing the model's economic environment.[3] Recently, Straub and Werning (2020) have demonstrated that the Chamley-Judd result does not hold even in the original models used to derive it by Judd (1985) and Chamley (1986).[4]

Within the voluminous literature, Aghion et al. (2013) is the first attempt to introduce R&D-based growth to the debate on the Chamley-Judd result. It is therefore worthwhile discussing our contribution and the main differences between our paper and Aghion et al. (2013). First, and most importantly, our study links optimal taxation to a detailed innovation process and monopolistic markup. Specifically, we examine how the optimal capital income tax responds to various R&D externalities by integrating the models in Aghion et al. (2013) and Jones and Williams (2000). The empirical literature on R&D has identified various externalities such as the "stepping on toes effect" and the "standing on shoulders effect" that are relevant to the innovation process.[5] Our study thus extends the Aghion et al. (2013) analysis by incorporating these externalities into the debate on optimal capital taxation. Second, by following the specification of Jones and Williams (2000), our model is free of the "scale effect" which is often not observed in reality (Jones, 1995).[6] Third, Aghion et al. (2013) find that a positive optimal capital income tax can be the case when the government spending-to-output ratio exceeds 38%, which is much larger than the empirical value. In

[3] The majority of this literature obtains a positive optimal capital income tax; see, e.g., Chari and Kehoe (1990), Stokey (1991), Aiyagari (1995), Chamley (2001), Erosa and Gervais (2002), Cozzi (2004), Domeij (2005), Conesa et al. (2009), Farhi and Werning (2010, 2013), Aghion et al. (2013), Chen and Lu (2013), Piketty and Saez (2013), Long and Pelloni (2017), Chang and Lai (2020), and Chen et al. (2021). Alternatively, a few studies overturn the Chamley-Judd result by proposing a negative optimal capital tax; see, e.g., Judd (1997, 2002), Coto-Martínez et al. (2007), and Petrucci (2015).

[4] Lansing (1999) obtains a similar conclusion to Straub and Werning (2020) but his paper is only limited to a special case in which the intertemporal elasticity of substitution is unity.

[5] See Neves and Sequeira (2018) and Sequeira and Neves (2020) for useful surveys of this literature.

[6] The earlier R&D-based growth models (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992) have a feature that changes in the size of an economy's population affect the long-run growth rate. Jones (1995) argues that such a "scale effect" is not supported by the empirical evidence.

our analysis, by contrast, the optimal capital income tax is positive even if the government spending ratio is reasonably small (around 17%). Finally, Aghion et al. (2013) consider a lab-equipment innovation process (i.e., R&D uses final goods as inputs) with Schumpeterian-type quality-improving R&D. Our model complements their analysis by considering a knowledge-driven innovation process (i.e., R&D uses labor as an input) with Romer-type expanding-variety R&D.

Long and Pelloni (2017) also find a sizable positive optimal capital tax in a standard expanding-variety R&D model à la Romer (1990). In contrast to previous literature on the Chamley-Judd result, in their analysis the role of physical capital is dismissed and the capital tax is imposed on the return of financial assets related to R&D investment. Moreover, their analysis has not examined the linkage between optimal taxation and various R&D externalities.

Our paper is also related to the literature on the "new dynamic public finance" approach developed by Golosov et al. (2003) who extend the static Mirrlees (1971) framework to dynamic settings. This literature considers a different environment from Judd (1985) and Chamley (1986) by emphasizing the roles of (aggregate and idiosyncratic) uncertainty and dynamic tax schemes, and finds that it is Pareto optimal to have positive capital taxes. The basic logic is that once capital is accumulated, it is sunk, and taxing capital is no longer distortionary. Thus it may become preferable for the government to deviate from the prescribed sequence of taxes by taxing capital.[7]

Finally, our paper is related to a group of studies that examine the effects of factor taxes in R&D-based growth models. Zeng and Zhang (2002) examine the long-run growth effects of various taxes including the capital, labor, and consumption tax. Scrimgeour (2015) examines the effects of reforming taxes on government revenues and welfare. Iwaisako (2016) explores

[7] See Golosov et al. (2006) for a useful reader's guide, and see Golosov and Tsyvinski (2015) for the recent development of this literature.

the effects of patent protection on optimal corporate income and consumption taxes. These papers, however, do not focus on the normative analysis of optimal capital taxation.

The rest of the paper proceeds as follows. In Section 2 we describe the R&D-based growth model featuring creative destruction and various types of R&D externalities elucidated by Jones and Williams (2000). In Section 3 we analyze how capital tax changes affect the economy in the long run. In Section 4 we quantify the optimal capital income tax rate and examine how its value depends on various R&D externalities. In Section 5 we provide several extensions to check the robustness of our results. Section 6 concludes this study.

2 The model

Our framework builds on the scale-invariant R&D-based growth model in Jones and Williams (2000). The main novelty of the Jones-Williams model is that it introduces a variety of R&D externalities into the original variety-expanding R&D-based growth model in Romer (1990). In this paper, we extend their model by incorporating (i) an elastic labor supply and (ii) factor income taxes, namely, capital and labor income taxes. To conserve space, the familiar components of the Romer variety-expanding model will be briefly described, while new features will be described in more detail.

2.1 Household

We consider a continuous-time economy that is inhabited by a representative household. At time t, the population size of the household is N_t, which grows at an exogenous rate n. Each member of the household is endowed with one unit of time that can be used to supply labor to a competitive market or enjoy leisure. The lifetime utility function of the representative household is given as:[8]

$$U = \int_0^\infty e^{-\beta t} \left[\ln c_t + \chi \ln(1 - l_t) \right] \, dt, \tag{1}$$

where c_t is per capita consumption and l_t is the supply of labor per capita. The parameters $\beta > 0$ and $\chi \geq 0$ denote, respectively, the subjective rate of time preference and leisure preference. The representative household maximizes (1) subject to the following budget constraint:

$$\dot{k}_t + \dot{e}_t = [(1 - \tau_K) r_{K,t} - n - \delta] k_t + (r_{e,t} - n) \, e_t + (1 - \tau_{L,t}) w_t l_t - c_t, \tag{2}$$

where a dot hereafter denotes the derivative with respect to time, k_t is physical capital per capita, $\delta > 0$ is the physical capital depreciation rate, e_t is the value of equity shares of R&D owned by each member, $r_{K,t}$ is the capital rental rate, $r_{e,t}$ is the rate of dividend, and w_t is the wage rate. The policy parameters τ_K and $\tau_{L,t}$ are respectively the capital and labor income tax rate.[9]

Solving the dynamic optimization problem yields the following first-order conditions:

$$\frac{1}{c_t} = q_t, \tag{3}$$

$$(1 - \tau_{L,t}) w_t (1 - l_t) = \chi c_t, \tag{4}$$

$$r_{e,t} = (1 - \tau_K) r_{K,t} - \delta. \tag{5}$$

where q_t is the Hamiltonian co-state variable on (2). Equations (3) and (4) are respectively the optimality conditions for consumption and labor supply, and (5) is a no-arbitrage condi-

[8] Here we assume that household welfare depends on per capita utility. See, e.g., Chu and Cozzi (2014) for a similar specification.

[9] We drop the subscript t for τ_K. In line with the literature, we consider τ_K as a time-independent policy parameter. As pointed out by Aghion et al. (2013), analyzing complicated time-dependent policies in the Ramsey framework is neither plausible nor empirically relevant.

tion which states that the net returns on physical capital and equity shares must be equalized. We denote the common net return on both assets as r_t (i.e., $r_t = r_{e,t} = (1-\tau_K)r_{K,t} - \delta$). The typical Keynes-Ramsey rule is:

$$\frac{\dot{c}_t}{c_t} = r_t - n - \beta. \tag{6}$$

2.2 The final-good sector

A perfectly-competitive final-good sector produces a single final output Y_t (treated as the numéraire) by using labor and a continuum of intermediate capital goods, according to the CES technology:

$$Y_t = L_{Y,t}^{1-\alpha}\left(\int_0^{A_t} x_t^{\alpha\rho}(i)di\right)^{\frac{1}{\rho}}, \alpha \in (0,1),\ \rho \in (0,\frac{1}{\alpha}), \tag{7}$$

where $L_{Y,t}$ is the labor input employed in final goods production, $x_t(i)$ is the i-th intermediate capital good, and A_t is the number of varieties of the intermediate goods.

Profit maximization yields the following conditional demand functions for the labor input and intermediate goods:

$$w_t = (1-\alpha)\frac{Y_t}{L_{Y,t}}, \tag{8}$$

$$p_t(i) = \alpha L_{Y,t}^{1-\alpha}\left(\int_0^{A_t} x_t^{\alpha\rho}(i)di\right)^{\frac{1}{\rho}-1} x_t^{\alpha\rho-1}(i), \tag{9}$$

where $p_t(i)$ is the price of the i-th intermediate good.

2.3 The intermediate-good sector

Each intermediate good is produced by a monopolistic producer that owns a perpetually protected patent for that good. The producer uses one unit of physical capital to produce one unit of intermediate goods; that is, the production function is $x_t(i) = v_t(i)$, where $v_t(i)$ denotes the capital input employed by monopolistic intermediate firm i. Accordingly, the profit of intermediate goods firm i is:

$$\pi_{x,t}(i) = p_t(i)x_t(i) - r_{K,t}v_t(i). \tag{10}$$

Let $\eta_t(i)$ denote the gross markup that the i-th intermediate firm can charge over its marginal cost; that is:

$$p_t(i) = \eta_t(i)r_{K,t}. \tag{11}$$

Then, the profit of the i-th intermediate firm can be obtained as:

$$\pi_{x,t}(i) = \frac{\eta_t(i) - 1}{\eta_t(i)}\alpha\frac{Y_t}{A_t}. \tag{12}$$

In subsection 2.5, we will elucidate how $\eta_t(i)$ is determined.

2.4 The R&D sector

R&D creates new varieties of intermediate goods for final-good production. The production technology we adopt incorporates the knowledge-driven specification of Romer (1990) and Jones (1995), i.e., innovation using the labor input (scientists and engineers), with the Jones and Williams (2000)'s specification which features fruitful R&D externalities:

$$(1 + \psi)\dot{A}_t = \Omega_t L_{A,t} \tag{13}$$

where $L_{A,t}$ is the labor input used in the R&D sector, and Ω_t is the productivity of R&D which the innovators take as given. The parameter $\psi \geq 0$ represents the size of the innovation clusters, which we will detail below.

We follow Jones (1995) to specify that the productivity of R&D takes the following functional form:

$$\Omega_t = L_{A,t}^{\lambda-1} A_t^{\phi}. \tag{14}$$

Eqatioans (13) and (14) contain three parameters λ, ϕ and ψ. These parameters capture salient features of the R&D process, as proposed by Jones and Williams (1998).

First, the parameter $\lambda \in (0, 1]$ reflects a (negative) duplication externality or a congestion effect of R&D. It implies that the social marginal product of research labor can be less than the private marginal product. This may happen because of, for example, a patent race, or if two researchers accidentally work out a similar idea. Jones and Williams (1998) refer to this negative duplication externality as the *stepping on toes effect.* Notice that this effect is stronger with a smaller λ, and it vanishes when $\lambda = 1$.

Second, the parameter $\phi \in (0, 1)$ reflects a (positive) knowledge spillover effect due to the fact that richer existing ideas are helpful to the development of new ideas. A higher ϕ means that the spillover effect is greater. In his pioneering article, Romer (1990) specifies $\phi = 1$; however, Jones (1995) argues that $\phi = 1$ exhibits a scale effect which is inconsistent with the empirical evidence. We follow Jones (1995) and assume that $\phi < 1$ in order to remove this scale effect. The knowledge spillover effect is dubbed by Jones and Williams (1998) as the *standing on shoulders effect.*

Finally, the parameter $\psi \geq 0$ denotes the size of the innovation clusters, which captures the concept of creative destruction formalized in the Schumpeterian growth model developed by Aghion and Howitt (1992). The basic idea is that innovations must come together in clusters, some of which are new, while others simply build on old fashions. More specifically, suppose that an innovation cluster, which contains $(1 + \psi)$ varieties, has been invented. Out of these $(1 + \psi)$ varieties, only one unit of variety is entirely new and thus increases the mass of the variety of intermediate goods. The remaining portion, of size ψ, simply replaces the old versions. This portion captures the spirit of creative destruction since new versions are created with the elimination of old versions. However this part does not contribute to an increase in existing varieties. In other words, for $(1+\psi)$ intermediate goods invented, the actual augmented variety is 1, while there are ψ repackaged varieties.

Given Ω_t, the R&D sector hires $L_{A,t}$ to create $(1+\psi)$ varieties. Thus, the profit function is $\pi_{A,t} = P_{A,t}(1+\psi)\dot{A}_t - w_t L_{A,t}$ where $P_{A,t}$ denotes the market value of a new variety. By assuming free entry in the R&D sector, we can obtain:

$$P_{A,t} = \frac{s_t}{1-s_t}\frac{(1-\alpha)Y_t}{(1+\psi)\dot{A}_t}, \tag{15}$$

where $s_t \equiv L_{A,t}/L_t$ is the ratio of research labor to total labor supply L_t. Moreover, the no-arbitrage condition for the value of a variety is:

$$r_t P_{A,t} = \pi_{x,t} + \dot{P}_{A,t} - \psi\frac{\dot{A}_t}{A_t}P_{A,t}. \tag{16}$$

In the absence of creative destruction ($\psi = 0$), the familiar no-arbitrage condition reports that, for each variety, the return on the equity shares $r_t P_{A,t}$ will be equal to the sum of the flow of the monopolistic profit $\pi_{x,t}$ plus the capital gain or loss $\dot{P}_{A,t}$. When creative destruction is present, existing goods are replaced. Accompanied by $\dot{A}_t$ new varieties being invented, the amount of $\psi\dot{A}_t$ existing varieties will be replaced. Therefore, for each variety, the expected probability of being replaced is $\psi\dot{A}_t/A_t$, which gives rise to the expected capital loss expressed by the last term in (16).

2.5 The monopolistic markup

This subsection explains how the monopolistic markup $\eta_t(i)$ is determined. As identified by Jones and Williams (2000), there are two scenarios in which the markup is decided. The first is the "unconstrained" case. In this case, the monopolistic intermediate firm freely sets the price by maximizing (10) subject to the production function $x_t(i) = v_t(i)$ and (9), which yields the pricing rule $p_t(i) = \frac{1}{\rho\alpha} r_{K,t}$. We refer to $\frac{1}{\rho\alpha}$ as the "unconstrained" markup. The second case is the "constrained" case, which may occur if the new designs are linked together in the *innovation cluster*. Specifically, a larger size of innovation clusters ψ serves as a constraint that controls the magnitude of the monopolistic markup. The intuition underlying this idea requires a more detailed explanation. Consider that the current number of varieties is A_t. Now an innovation cluster with size $(1+\psi)$ is developed. This increases the mass of varieties to A_t+1; at the same time it also replaces old-version varieties by ψ units. Subsequently, the final-good firm faces two choices. It can either adopt the new innovation cluster and then use A_t+1 intermediate goods priced at a markup, or part with the new innovation cluster and still use A_t intermediate goods in the production process. If the final-good firm chooses the latter, since ψ varieties have now been displaced, the final-good firm only needs to purchase $A_t - \psi$ units of intermediate goods at a markup price, while the other ψ units of displaced intermediate goods can be purchased at a lower (competitive) price. When the size of an innovation cluster is high (a large value of ψ), the final-good firm will not tend to adopt the new innovation cluster because sticking to old clusters is cheaper. As a result, the intermediate-good firms have to set a lower price so as to attract the final-good firm to adopt the new innovation cluster. This "adoption constraint" explains why an increase in the size of the innovation clusters reduces the markup.

In an appendix, Jones and Williams (2000) demonstrate that the constrained markup is negatively related to both the size of the innovation clusters and the elasticity of substitution between capital goods. Specifically, they demonstrate that, in order to attract the final-good firm to adopt the new innovation cluster, the intermediate-good firms cannot set a markup that is higher than $[(1+\psi)/\psi]^{1/\rho\alpha-1}$. A profit-maximizing firm thus always tends to set the highest price $p_t(i) = [(1+\psi)/\psi]^{1/\rho\alpha-1} r_{K,t}$. We refer to $[(1+\psi)/\psi]^{1/\rho\alpha-1}$ as the "unconstrained" markup. By combining the constrained markup pricing with the unconstrained markup pricing rule mentioned earlier (i.e., $p_t(i) = \frac{1}{\rho\alpha} r_{K,t}$), we can conclude that the equilibrium markup is:

$$\eta_t(i) = \min\left\{ \frac{1}{\rho\alpha} \, , \, \left(\frac{1+\psi}{\psi}\right)^{\frac{1}{\rho\alpha}-1} \right\}, \tag{17}$$

which is independent of i and t. Combining (10) and (17) implies that all intermediate-good firms are symmetric. Hence, the notation i can be dropped from now on.

2.6 The government and aggregation

The government collects capital income taxes and labor income taxes to finance its public spending. The balanced budget constraint faced by the government is:

$$N_t(\tau_K r_{K,t} k_t + \tau_{L,t} w_t l_t) = G_t, \tag{18}$$

where G_t is the total government spending. We assume that government spending is a fixed proportion of final output, i.e., $G_t = \zeta Y_t$, where $\zeta \in (0, 1)$ is the ratio of government spending to output. As in Conesa et al. (2009), Aghion et al. (2013) and Long and Pelloni (2017), equation (18) puts aside the role of government debt when examining the optimal factor taxes. That is, we mainly focus on the trade-off between the capital and labor income tax.

Now let us define the aggregate capital stock as $K_t = N_t k_t$, aggregate consumption $C_t = N_t c_t$, and total labor supply $L_t = N_t l_t$. After some derivations, we can obtain the following resource constraint in the economy: $\dot{K}_t = Y_t - C_t - G_t - \delta K_t$.

2.7 The decentralized equilibrium

The decentralized equilibrium in this economy is an infinite sequence of allocations $\{C_t,\ K_t,\ A_t,\ Y_t,\ L_t,\ L_{Y,t},\ L_{A,t},\ x_t,\ v_t\}_{t=0}^{\infty}$, prices $\{w_t,\ r_{K,t},\ r_t,\ p_t,\ P_{A,t}\}_{t=0}^{\infty}$, and policies $\{\tau_K,\ \tau_{L,t}\}$, such that at each instant of time:

a. households choose $\{c_t,\ k_t,\ e_t,\ l_t\}$ to maximize lifetime utility, (1), taking prices and policies as given;

b. competitive final-good firms choose $\{x_t,\ L_{Y,t}\}$ to maximize profit taking prices as given;

c. monopolistic intermediate firms $i \in [0, A_t]$ choose $\{v_t,\ p_t\}$ to maximize profit taking $r_{K,t}$ as given;

d. the R&D sector chooses $L_{A,t}$ to maximize profit taking $\{P_{A,t},\ w_t\}$ and the productivity Ω as given;

e. the labor market clears, i.e., $N_t l_t = L_{A,t} + L_{Y,t}$;

f. the capital market clears, i.e., $N_t k_t = A_t v_t$;

g. the stock market for variety clears, i.e., $N_t e_t = P_{A,t} A_t$

h. the resource constraint is satisfied, i.e., $\dot{K}_t = Y_t - C_t - G_t - \delta K_t$;

i. the government budget constraint is balanced, i.e., $N_t(\tau_K r_{K,t} k_t + \tau_{L,t} w_t l_t) = G_t$.

3 Balanced growth path

In this section, we explore the balanced growth path along which each variable grows at a constant rate (some can be zero). We denote the growth rate of any generic variable Z by

g_Z, and drop the time subscript to denote any variable in a steady state. The steady-state growth rates of varieties and output are given by (see Appendix A):

$$g_A = \frac{\lambda}{1-\phi}n, \quad g_Y = \frac{1}{1-\alpha}\left(\frac{1}{\rho} - \alpha\right) g_A + n. \tag{19a}$$

Moreover, in order to obtain stationary endogenous variables, it is necessary to define the following transformed variables:

$$\hat{k}_t \equiv \frac{K_t}{N_t^\sigma}, \quad \hat{c}_t \equiv \frac{C_t}{N_t^\sigma}, \quad \hat{y}_t \equiv \frac{Y_t}{N_t^\sigma}, \quad \hat{a}_t \equiv \frac{A_t}{N_t^{\lambda/(1-\phi)}}, \tag{19b}$$

where $\sigma \equiv 1 + \frac{(1/\rho-\alpha)\lambda}{(1-\alpha)(1-\phi)} > 0$ is a composite parameter. For ease of exposition, in line with Eicher and Turnovsky (2001), $\hat{k}$, $\hat{c}$, $\hat{y}$, and $\hat{a}$ are dubbed the scale-adjusted capital, consumption, output, and R&D varieties, respectively. Based on the transformed variables and the equilibrium defined in subsection 2.5, the economy in the steady state can be described by the following set of equations:

$$
\begin{aligned}
r &= (1-\tau_K)r_K - \delta = \beta + g_Y, && (20a)\\
s &= \frac{\frac{\eta-1}{\eta}\frac{\alpha}{1-\alpha}(1+\psi)g_A}{r - g_Y + \left(1+\frac{\eta-1}{\eta}\frac{\alpha}{1-\alpha}\right)(1+\psi)g_A}, && (20b)\\
\frac{\hat{k}}{\hat{y}} &= \frac{\alpha}{\eta r_K}, && (20c)\\
(1-\zeta)\frac{\hat{y}}{\hat{k}} &= \frac{\hat{c}}{\hat{k}} + g_Y + \delta, && (20d)\\
\hat{y} &= \hat{a}^{1/\rho-\alpha}\hat{k}^{\alpha}\left((1-s)l\right)^{1-\alpha}, && (20e)\\
g_A &= \frac{1}{1+\psi}\frac{(sl)^{\lambda}}{\hat{a}^{1-\phi}}, && (20f)\\
\frac{\chi l}{(1-l)} &= \frac{(1-\tau_L)(1-\alpha)}{(1-s)}\frac{\hat{y}}{\hat{c}}, && (20g)\\
\tau_L &= \frac{1-s}{1-\alpha}\left(\zeta - \tau_K\frac{\alpha}{\eta}\right), && (20h)
\end{aligned}
$$

in which eight endogenous variables r, s, $\hat{c}$, $\hat{k}$, $\hat{a}$, $\hat{y}$, l, τ_L are determined.

Of particular note, our main focus is on the examination of the capital tax. By holding the proportion of the government spending constant, an increase in the capital income tax will be coupled with a reduction in the labor income tax. Therefore, the literature on the Chamley-Judd result generally assumes that the labor income tax endogenously adjusts to balance the government budget. This approach has been dubbed as "tax shifting" or "tax swap" in the literature. Our analysis follows this standard approach in the literature.

3.1 Comparative statics analysis

In this subsection, we analyze the effects of capital taxation on the R&D share of labor s, the endogenous labor income tax rate, labor supply, and other scale-adjusted variables: $\hat{a}, \hat{k}, \hat{c}$, and $\hat{y}$.[10]

The long-run R&D labor share, s, is given by:

$$s = \frac{\frac{\eta-1}{\eta}\frac{\alpha}{1-\alpha}(1+\psi)g_A}{r - g_Y + \left(1 + \frac{\eta-1}{\eta}\frac{\alpha}{1-\alpha}\right)(1+\psi)g_A}. \tag{21a}$$

From (21a) we have the following proposition:

Proposition 1 *In the steady state, the R&D labor share is independent of the capital income tax rate.*

The intuition underlying Proposition 1 can be grasped as follows. The non-arbitrage condition between physical capital and R&D equity reported in (20a) requires that the return on physical capital be equal to the return on R&D equity. Given that the return on R&D equity, $r = \beta + \frac{1}{1-\alpha}\left(\frac{1}{\rho} - \alpha\right) g_A + n$, is independent of the capital tax rate, the capital income tax rate then does not affect the return on R&D equity and the R&D labor share. Therefore, our analysis does not rely on the channel that capital taxation generates a direct effect on the allocation of R&D and production labor. Instead, our analysis is based on the trade-off between labor supply and capital investment as in the standard Chamley-Judd setting.

From (20h), we have:

$$\tau_L = \frac{1-s}{1-\alpha}\left(\zeta - \tau_K \frac{\alpha}{\eta}\right), \tag{21b}$$

Based on (21b), we have:

$$\frac{\partial \tau_L}{\partial \tau_K} = -\frac{1-s}{\eta}\frac{\alpha}{1-\alpha} < 0. \tag{21c}$$

[10] We solve the dynamic system in Appendix B, and a detailed derivation of the comparative static analysis is presented in Appendix C.

The above equation shows that an increase in the capital income tax rate is coupled with a reduction in the labor income tax rate.

Given a constant capital income tax rate τ_K, labor supply in the steady state is given by:

$$l = 1 - \frac{\chi}{\chi + \frac{1}{[(1-\zeta)-(\delta+g_Y)\frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)}]}\frac{(1-\tau_L)(1-\alpha)}{(1-s)}}. \tag{22a}$$

It is straightforward from (22a) to infer the following result:

$$\frac{\partial l}{\partial \tau_K} = \frac{\alpha\beta(\frac{1-s}{1-\alpha})[1-\zeta+\frac{\eta-1}{\eta}\frac{\alpha(\delta+g_Y)}{\beta+(1+\psi)g_A}](1-l)l}{\eta(\beta+\delta+g_Y)(1-\tau_L)[1-\zeta-(\delta+g_Y)\frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)}]} > 0. \tag{22b}$$

Moreover, the scale-adjusted R&D varieties $\hat{a}$ is given by:

$$\hat{a} = \left[\frac{1}{(1+\psi)g_A}\right]^{1/(1-\phi)} (sl)^{\lambda/(1-\phi)}, \tag{23a}$$

where s and l are reported in (21a) and (22a). With $\partial s/\partial \tau_K = 0$, it is quite easy from (23a) to derive that:

$$\frac{\partial \hat{a}}{\partial \tau_K} = \frac{\lambda \hat{a}}{(1-\phi)l}\frac{\partial l}{\partial \tau_K} > 0. \tag{23b}$$

We summarize the above results in the following proposition.

Proposition 2 *Under an elastic labor supply ($\chi > 0$), a rise in the capital income tax rate boosts labor supply and (scale-adjusted) R&D varieties.*

The intuition underlying Proposition 2 can be explained as follows. In response to a rise in the capital income tax rate, the following effect emerges. Raising the capital tax rate reduces the labor income tax rate (see (21c)) and raises the after-tax wage income, thereby exerting a positive effect on labor supply. Therefore, a rise in the capital income tax rate is accompanied by an increase in labor supply. This in turn increases the labor input allocated to the R&D sector ($L_A = Nsl$). Then, as reported in (23a), given that scale-adjusted R&D varieties $\hat{a}$ is increasing in the R&D labor input sl, $\hat{a}$ will increase in response to a rise in τ_K.

We now examine the effect of the capital tax on output. From (20a), (20c), (20d), (20e), and (23a), we can infer that:

$$\hat{y} = \left[\frac{1}{(1+\psi)g_A}\right]^{\frac{1/\rho-\alpha}{(1-\alpha)(1-\phi)}} (sl)^{\frac{1/\rho-\alpha}{1-\alpha}\frac{\lambda}{1-\phi}} \left[\frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)}\right]^{\frac{\alpha}{1-\alpha}} (1-s)l, \tag{24a}$$

where

$$\frac{\partial \hat{y}}{\partial \tau_K} = \left[-\frac{\alpha}{(1-\alpha)(1-\tau_K)} + \frac{\sigma}{l}\frac{\partial l}{\partial \tau_K}\right]\hat{y} \gtrless 0. \tag{24b}$$

Thus, we have:

Proposition 3 *Under an inelastic labor supply ($\chi = 0$), a rise in the capital income tax rate lowers final output, whereas under an elastic labor supply ($\chi > 0$), the effect of the capital income tax on final output is uncertain.*

We explain the intuition as follows. As shown in (24b), two conflicting effects emerge following a rise in the capital income tax rate. First, a rise in the capital income tax rate shrinks capital investment, which in turn generates a negative impact on output. Second, a rise in the capital income tax rate is accompanied by a fall in the labor income tax rate, which motivates the household to provide more labor supply. This increase in labor supply implies that more labor input is available for the R&D sector and in turn boosts R&D varieties, thereby contributing to a positive effect on output. If labor supply is exogenous ($\chi = 0$), the second positive effect is absent ($\partial l/\partial \tau_K = 0$), and thus a higher capital income tax rate lowers output. However, if labor supply is endogenous ($\chi > 0$), the two opposing effects are present, and the output effect of capital income taxation depends upon the relative strength between these two effects.

Finally, we examine the effects of the capital tax on capital and consumption. From (20a), (20c), and (20d), we have:

$$
\begin{aligned}
\hat{k} &= \frac{(1-\tau_K)\Phi}{(\delta+g_Y)}\hat{y}, && (25a)\\
\hat{c} &= [(1-\zeta)-(1-\tau_K)\Phi]\hat{y}, && (25b)
\end{aligned}
$$

where $\Phi \equiv \frac{\alpha(\delta+g_Y)}{\eta(\beta+\delta+g_Y)}$ is a composite parameter. Based on (25a) and (25b), the effects of τ_K on $\hat{k}$ and $\hat{c}$ can be expressed as:

$$
\begin{aligned}
\frac{\partial \hat{k}}{\partial \tau_K} &= -\frac{\Phi}{(\delta+g_Y)}\hat{y} + \frac{(1-\tau_K)\Phi}{(\delta+g_Y)}\frac{\hat{y}}{\partial \tau_K} && (26a)\\
&= [\sigma\frac{\partial l}{l\partial \tau_K} - \frac{1}{(1-\alpha)(1-\tau_K)}]\frac{(1-\tau_K)\Phi}{(\delta+g_Y)}\hat{y}\gtrless 0,\\
\frac{\partial \hat{c}}{\partial \tau_K} &= \Phi\hat{y} + [(1-\zeta)-(1-\tau_K)\Phi]\frac{\partial \hat{y}}{\partial \tau_K} && (26b)\\
&= \{\Phi + [(1-\zeta)-(1-\tau_K)\Phi][\sigma\frac{\partial l}{l\partial \tau_K} - \frac{\alpha}{(1-\alpha)(1-\tau_K)}]\}\hat{y}\gtrless 0.
\end{aligned}
$$

We see from (26a) and (26b) that both effects of the capital tax on capital and consumption are ambiguous. It is clear in (25a) that capital income taxation affects scale-adjusted capital $\hat{k}$ through two channels. The first channel is the capital-output ratio $\hat{k}/\hat{y} = \frac{(1-\tau_K)\Phi}{(\delta+g_Y)}$, and the second channel is the level of scale-adjusted output $\hat{y}$. The first term after the first equality in (26a) indicates that the first channel definitely lowers the level of $\hat{k}$. Moreover, as shown in (24b), the second channel may either raise or lower the level of $\hat{k}$ since capital taxation leads to an ambiguous effect on $\hat{y}$. As a consequence, the net effect of capital taxation on the scale-adjusted capital stock $\hat{k}$ is still uncertain. Similarly, as indicated in (25b), capital income taxation also affects $\hat{c}$ through two channels. The first channel is the consumption-output ratio $\hat{c}/\hat{y} = [(1-\zeta) - (1-\tau_K)\Phi]$, and the second channel is the level of scale-adjusted output $\hat{y}$. The first channel definitely boosts the level of $\hat{c}$, while the second channel may either raise or lower the level of $\hat{c}$ since capital taxation leads to an ambiguous effect on $\hat{y}$. As a consequence, the net effect of capital taxation on scale-adjusted consumption $\hat{c}$ remains ambiguous.

4 Quantitative results

In this section, we simulate the welfare effects of capital taxation and compute the optimal capital tax rate by performing a quantitative analysis.[11] We calibrate the parameters of our theoretical model based on US data to quantify the optimal capital tax. Then we explore how the optimal capital tax responds to important parameters that feature R&D externalities and the government size.[12]

By dropping the exogenous terms, the life-time utility of the representative household reported in (1) can be expressed as:

$$U = \int_0^\infty e^{-\beta t} \left[\ln \hat{c}_t + \chi \ln(1-l_t)\right] \, dt, \tag{27}$$

in which $\hat{c}_t$ and l_t are functions of τ_K. The government chooses the capital income tax rate τ_K to maximize (27) while balancing the budget, (18), by using the labor tax.

[11] We focus on the optimal capital tax that maximizes the steady-state welfare, to be consistent with Jones and Williams (2000) and Aghion et al. (2013).

[12] We start from the same initial steady state when we vary the value of each parameter.

4.1 Calibration

To carry out a numerical analysis, we first choose a baseline parameterization, as reported in Table 1. Our model has eleven parameter values to be assigned. These parameters are either set to a commonly used value in the existing literature or calibrated to match some empirical moments in the US economy. We now describe each of them in detail. In line with Andolfatto et al. (2008) and Chu and Cozzi (2018), the labor income share $1 - \alpha$ and the discount rate β are set to standard values 0.4 and 0.05, respectively. The population growth rate n is set to 0.011 as used by Conesa et al. (2009). The physical capital depreciation rate is set to 0.0318 so that the initial capital-output ratio is 2.5 as in Lucas (1990). The initial capital tax rate τ_K is set to 0.3 based on the average US effective tax rate estimated by Carey and Tchilingurian (2000). A similar value of the capital income tax rate has been adopted in Domeij (2005) and Chen and Lu (2013). As for the government size (the ratio of government spending to output), data for the US indicate that this is around 20% (Gali, 1994), and has slightly increased in recent years. We therefore set ζ to be 0.22, which is the average level during 2001-2013, to reflect its increasing trend. The parameter for leisure preference χ is chosen as 1.619 to make hours worked one third of total hours (Jermann and Quadrini, 2012).

[Table 1 here]

Now we deal with the parameterization regarding the R&D process. First, Sequeira and Neves (2020) review the empirical literature on R&D and provide a meta-analysis concerning the value of the stepping on toes effect λ, and find that the average size is around 0.2. This value is also within the reasonable range estimated by Bloom et al. (2013). We follow these empirical studies to set $\lambda = 0.2$ as our benchmark. The substitution parameter ρ is closely related to the markup of the intermediate firms. We set ρ to be 2.2727 such that, given $1 - \alpha$, the (unconstrained) markup in our economy is 1.1, which lies within the reasonable range estimated for US industries (e.g., Laitner and Stolyarov, 2004; Yang, 2018). Next, we use the output growth rate to calibrate the extent of the standing on shoulders effect ϕ. In our model we have:

$$g_Y = \frac{1}{1-\alpha}\left(\frac{1}{\rho} - \alpha\right) g_A + n. \tag{28}$$

Given that $g_A = \lambda n/(1-\phi)$ and that we have already assigned values to $1-\alpha$, ρ, n and λ, we can then choose ϕ to target the empirical level of the output growth rate in the US. Specifically, we set $\phi = 0.9837$ as our baseline value such that the output growth rate is 2%. The value of ϕ that we choose is well supported by empirical studies (Ang and Madsen, 2015; Sequeira and Neves, 2018). Finally, as a benchmark we choose the size of the innovation cluster $\psi = 0.25$ by following Comin (2004). In this case the markup is not bound by the adoption constraint. If the value of ψ is large, the markup will then be constrained by this parameter. Under our baseline parameter values, the consumption-GDP ratio is around 69% and the R&D-GDP ratio is around 3%. These values fit the US data.

4.2 The optimal capital tax rate

Figure 1 plots the relationship between the level of welfare cost (in terms of the percentage of consumption loss) and the rate of capital income tax, which exhibits an U-shaped relationship.[13] The optimal capital tax rate is positive, and its value is around 6.6%. Thus the Chamley-Judd result of zero capital tax does not hold in our R&D-based growth model.

[Figure 1 here]

The intuition underlying this result is as follows. Given that the government is limited to capital and labor taxation to finance a fix amount of the government expenditure, not taxing

[13] As in Jones and Williams (2000), we focus on the optimal capital tax rate that maximizes steady-state welfare. If we instead consider welfare including the transitional dynamics, the optimal capital tax rate would be higher. See Chen et al. (2019) for the level of optimal capital tax rate for transitional welfare.

capital income implies that the labor income must be taxed at a higher rate. Although a zero capital tax efficiently leaves the capital market undistorted, a high labor tax distorts the labor market severely by decreasing the after-tax wage income and in turn reduces total labor supply. As a consequence, there is less labor devoted to the production in the R&D sector, which then results in fewer equilibrium varieties for the final-good production, and ultimately depresses the level of consumption and welfare.

Of particular note, to highlight the roles of R&D externalities, our study follows the original Chamley-Judd framework in assuming that there are is no uncertainty and heterogeneity. This simplicity implies that the objective of taxation is purely on the efficiency aspect, and absent from the insurance and redistribution purposes. If idiosyncratic risks are present, taxation serves the purpose of an insurance by redistributing income from high-income households to low-income households, which would give rise to a higher optimal capital tax rate. For example, Imrohoroglu (1998) and Conesa et al. (2009) introduce uninsurable idiosyncratic income risks as an important rationale for positive capital taxation. In Conesa et al. (2009) the suggested optimal capital tax rate goes up to 36%, which is much higher than our result. This means that introducing idiosyncratic uncertainty into our present model would make the optimal capital income tax even more likely to be positive. Future research could be extended in this direction.

Previous studies on optimal capital taxation in endogenous growth models (Aghion et al., 2013; Long and Pelloni, 2017) have pointed out that the welfare-maximizing capital tax rate is not equal to the growth-maximizing capital tax rate. In our semi-endogenous growth model, the steady-state growth rate is independent of the tax parameters. However, it is equally important, and of policy relevance, to examine the capital tax rate that maximizes (scale-adjusted) final output in this model, and compare it to the optimal capital tax rate. Under the benchmark, we find that the output-maximizing capital tax rate is 4.2%, which is lower than the welfare-maximizing capital tax rate of 6.6%. The implication of this result is that taxing capital can be good for welfare even if it is harmful for GDP. This result echoes that of Long and Pelloni (2017) who find that raising the capital tax may improve welfare while reducing growth.

Although Figure 1 suggests a positive optimal capital tax, we should note that this result is obtained under our benchmark parameters, and it may change when the innovation process exhibits different degrees of R&D externalities. Thus, our goal is not to conclude that it is always right to tax capital, but to highlight that in achieving the social optimum, it is necessary to balance both distortions in the capital and labor markets. In view of this, an extreme case of the zero capital tax is often suboptimal. More importantly, we make an attempt to give guidance on which R&D mechanisms are at play in influencing the optimal capital tax, which we will show in the next subsection.

4.3 Policy implications of R&D externalities

In this subsection, we investigate how the optimal capital tax responds to relevant parameters, in particular those related to the innovation process. More importantly, we shed some light on the roles of R&D externalities in the design of optimal tax policies. In what follows, we propose some relevant parameters that need to be considered by the policy-makers. The results are depicted in Figures 2 to 6.

[Figures 2 and 3 here]

First, Figures 2 and 3 show that the optimal capital tax rate is increasing in λ (the stepping on toes effect) and ϕ (the standing on shoulders effect). With sufficiently small values of λ and ϕ, the optimal capital income tax is negative. Notice that a higher λ implies that the negative duplication externality is small, and a higher ϕ means that the positive spillover effect of R&D is relatively strong. Both cases imply a similar circumstance in which the innovation process is more productive, and in which underinvestment in R&D is more likely. Under such a situation, the welfare cost of depressing innovation by raising the labor income tax is larger. Therefore, the government is inclined to increase the capital tax while reducing the labor tax.

[Figure 4 here]

Second, Figure 4 shows that the optimal capital income tax and the substitution parameter ρ exhibit an inverted-U shaped relationship. A lower ρ is associated with a higher monopolistic markup η, regardless of whether the adoption constraint is binding or not. The substitution parameter mainly affects the optimal capital tax in three different ways. First, when η is large (when ρ is small), the degree of the intermediate firms' monopoly power is strong. To correct this distortion, the government tends to subsidize capital to offset the gaps between price and the marginal cost; see Judd (1997, 2002). Second, when η is large (when ρ is small), the private value of inventions increases. As a result, equilibrium R&D increases, which in turn makes R&D overinvestment more likely. Therefore, the government tends to raise the tax on labor because R&D uses labor in our model. These two effects indicate that the optimal capital tax should be decreasing in the markup as in previous studies. Third, a small ρ amplifies the productivity of varieties in final-good production and thus amplifies the effect of g_A on g_Y (see (28)). In this case, the government is inclined to subsidize labor by taxing capital since the R&D sector uses labor. This last effect indicates that the optimal capital tax rate is decreasing in the elasticity of substitution between intermediate goods (or increasing in the markup). Figure 4 shows that the first two effects dominate when ρ is small and the third effect dominants when ρ becomes sufficiently large. Thus the optimal capital tax reverses as ρ exceeds a threshold value.

[Figure 5 here]

Third, Figure 5 shows that the optimal capital tax decreases in response to a rise in the size of the innovation cluster (creative destruction). To explain the intuition, we first distinguish three effects that creative destruction may have on the incentive to engage in R&D. The first positive effect comes from the R&D firm being able to earn profits even for those of its products that do not really increase the variety of intermediate goods (note that $\pi_{A,t} = P_{A,t}(1+\psi)\dot{A}_t - w_t L_{A,t}$).[14] This is referred to as the "carrot" by Jones and Williams (2000). The second negative effect arises, as exhibited in (15), from a higher ψ that decreases the equilibrium price of the products in the presence of free entry, even though it increases the products sold by the R&D firm. The third negative effect is associated with the no-arbitrage condition for the value of a variety, which is displayed in (16). Due to creative destruction, existing goods have a probability of being replaced by new goods, and this probability increases with the degree of creative destruction. Therefore, creative destruction increases the expected capital loss in terms of the return on the equity shares, and in turn reduces the incentive to engage in R&D. Jones and Williams (2000) dub this effect as the "stick". Figure 5 shows that the positive carrot effect dominates the negative stick effect, which is in consistent with the result in Jones and Williams (2000). As a consequence, a higher ψ stimulates R&D, which makes R&D overinvestment more likely. Hence the government should decrease the capital tax and increase the labor tax.

[Figure 6 here]

Finally, the optimal capital tax is increasing in the government spending ratio ζ, as shown in Figure 6. This result is consistent with Aghion et al. (2013) and Lu and Chen (2015).[15] When the need for public expenditure is sufficiently small, the government can collect labor tax revenues to finance the government spending and also to subsidize capital. Note that in

[14] The R&D firm can earn profits from its whole products $(1+\psi)\dot{A}$, in which $\psi\dot{A}$ does not contribute to the increase of varieties.

[15] Lu and Chen (2015) show that in an exogenous growth model with a given share of government expenditure in output, the optimal capital income tax is positive and increasing with the the share of government expenditure. The intuition is that capital accumulation reduces the discounted net marginal product of next period's capital by way of increasing government expenditure. Thus, the government should tax capital to correct this distortion.

this case the monopoly effect dominates the R&D effect so that the optimal capital tax rate becomes negative. As the size of government expenditure increases, it is not promising to rely solely on raising the labor tax, because the distortion to the R&D sector would be too strong. In this case, it becomes optimal to shift some of the tax burden to capital.

As we have noted earlier, our result of a positive optimal capital income tax is obtained under the benchmark parameters. Before ending this section, it is worthwhile to briefly discuss how plausible the above parameters fall into the range that implies a negative optimal capital tax rate. First, the optimal capital tax rate becomes negative if $\lambda < 0.14$, namely when the stepping on toes effect is very strong. Second, the optimal capital tax rate becomes negative if the standing on shoulders effect is smaller, i.e., $\phi < 0.976$. According to the estimation by Ang and Madsen (2015), the value of ϕ is close to one. Moreover, for the first-generation R&D-based growth models à la Romer (1990), $\phi = 1$, so that our result of a positive optimal capital income tax always holds. Third, for the substitution parameter ρ, the threshold values that will result in a negative optimal capital tax is $\rho > 2.37$ and $\rho < 0.58$. Both values correspond to implausible monopolistic markups, i.e., $\eta < 1.05$ and $\eta > 4.31$, respectively. Fourth, the optimal capital tax rate is positive for the whole possible values of the size of the innovation cluster ψ. Finally, the optimal capital tax rate is negative if the government spending ratio is less than 16.5%, i.e., $\zeta < 0.165$. This threshold value is much smaller than that in Aghion et al. (2013), in which the government spending ratio required for a positive optimal capital tax rate is around 38%.

5 Extensions and discussions

In this section, we perform several extensions to the benchmark model in order to explore the robustness of our results. We expand the policy instruments available to the government in subsection 5.1 (a corporate profit tax) and subsection 5.2 (progressivity of labor income tax).

We also adopt various specifications regarding the labor-leisure preference. In subsection 5.3 we consider an iso-elastic labor supply function to explore the role of labor supply elasticity. In subsection 5.4 we target different labor-leisure allocations and examine its effect on the optimal capital tax, and we discuss the application of our result to the optimal capital tax in the EU.

5.1 Corporate profit tax

In our benchmark model, we assume that the government only has access to the capital and labor income tax. In this subsection, we extend the tax regime to include a corporate profit tax. Let $\tau_{\pi,t}$ denote the corporate profit tax on the intermediate-good firms' profit. Accordingly, the after-tax profit of the intermediate-good firms is

$$(1-\tau_{\pi,t})\pi_{x,t}(i) = (1-\tau_{\pi,t})[p_t(i)x_t(i) - r_{K,t}v_t(i)]. \tag{29}$$

Moreover, with symmetric intermediate-good firms, the government budget constraint is modified as:

$$N_t(\tau_K r_{K,t}k_t + \tau_{L,t}w_t l_t) + A_t\tau_{\pi,t}\pi_{x,t} = G_t. \tag{30}$$

With this new tax regime, we conduct two policy experiments. In the first experiment, we engage in tax shifting between the capital income tax and labor income tax with the presence of the corporate tax. That is, we treat the corporate profit tax as an exogenous variable, i.e., $\tau_{\pi,t} = \tau_\pi$. We then examine how the optimal capital tax rate responds to the corporate profit tax. The result is shown in Figure 7.

[Figure 7 here]

We see in Figure 7 that the optimal capital tax rate is positive and decreasing in the corporate profit tax. Intuitively, there are two conflicting effects of the corporate profit tax on the optimal capital tax. On the one hand, the corporate profit tax decreases the after-tax profit of the intermediate-good firms. This depresses the incentives for R&D and causes underinvestment in R&D. Thus, the government should raise the capital income tax while reducing the labor income tax. On the other hand, when the corporate tax rate is higher, the government can collect more profit tax revenues, which means that the tax revenues required from taxing capital and labor income are smaller. This effect therefore acts like the effect of a smaller government size that we have discussed using Figure 6. As such, the government tends to reduce the capital income tax. Figure 7 shows that the latter effect outweighs the former, so that a higher corporate profit tax is associated with a lower optimal capital tax.

[**Figure 8 here**]

In the second experiment, we consider tax shifting between the capital income tax and the corporate profit tax. We assume that the labor income tax is exogenous, and set its benchmark value to 17.6%. Figure 8 shows how the optimal capital tax rate responds to the labor income tax. In this experiment, the government faces a trade-off between taxing capital income and firms' profits. While taxing capital income creates a dynamic inefficiency for capital accumulation, taxing profits depresses the incentives for innovation. Under the benchmark parameters, we find that the optimal capital tax rate is 31%, and decreasing with the labor income tax rate. This result is quite intuitive. When the labor income tax rate is higher, the government does not need to tax the capital income or profits so much.

5.2 Tax progressivity

In this subsection, we extend the tax regime to consider a progressive labor income tax, which is more realistic in many countries including the US. For tractability, we follow the specification in Guo and Lansing (1998) by assuming that:

$$\tau_{L,t} = 1 - \mu_t \left(\frac{\bar{I}_{L,t}}{I_{L,t}(j)} \right)^{\varepsilon}. \tag{31}$$

In this specification, $\bar{I}_{L,t} = w_t \bar{l}_t$ is the average level of labor income per household with $\bar{l}_t$ being the average labor supply, $I_{L,t}(j) = w_t l_t(j)$ is the taxable labor income of household j, and μ_t determines the level of the labor income tax rate. Note that μ_t is an endogenous variable that adjusts to balance the government budget. The parameter $\varepsilon \in [0, 1)$ governs the degree of progressivity of the labor income tax. We see from (31) that if $\varepsilon > 0$, the household faces a tax rate that is increasing in its taxable labor income. A larger ε corresponds to a higher degree of progressivity because the tax rate is more sensitive to personal income. In the case where $\varepsilon = 0$, the household faces a flat tax rate equal to $1 - \mu_t$, and the model reverts to our basic model.

[Figure 9 here]

Figure 9 depicts how the optimal capital tax responds to the tax progressivity. Mattesini and Rossi (2012) compute the progressivity parameter of many countries over the period 1999-2009, and obtain a value of $\varepsilon = 0.18$ for the US. By applying this value to our model, we derive an optimal capital income tax rate of 11.2%. However, Chen and Guo (2013) point out that tax progressivity has recently exhibited a significant downward trend in the US, and suggest a smaller value of $\varepsilon = 0.063$. In Figure 9, this corresponds to an optimal capital income tax rate of around 7%.[16] To sum up, our result of a positive optimal capital tax is robust when tax progressivity is considered. Furthermore, the optimal capital tax rate

[16] In Mattesini and Rossi (2012) and Chen and Guo (2013), the progressivity parameter represents the degree of progressivity of total income, while in our model it represents the degree of progressivity of labor income. Given that labor-income taxation is more progressive than capital-income taxation in the US, this implies that our ε should be higher than their estimated values. Accordingly, the optimal capital tax rate could be even higher than the values proposed in the text.

is increasing in the progressivity of the labor income tax. The intuition is that when the labor income tax is more progressive, raising its tax rate results in higher distortion in the labor market. Therefore, the government tends to raise the capital income tax rate as it is relatively less distortive.

5.3 The elasticity of labor supply

Our benchmark model assumes a log-utility function, and thus dismisses the role played by the elasticity of labor supply in determining the trade-off between the capital and labor income tax. In this subsection, we extend the model to examine how the optimal capital tax rate varies with the elasticity of labor supply. To do so, we generalize the utility function to the iso-elastic labor supply function, which is given by:

$$U = \int_0^\infty e^{-\beta t} \left[\ln c_t + \chi \frac{(1 - l_t)^{1-\theta}}{1 - \theta} \right] dt, \tag{1'}$$

where $\theta > 0$ is the inverse of the Frisch elasticity of labor supply. We recalibrate the model. The benchmark value of θ is chosen as 0.833, which implies that the Frisch elasticity of labor supply is 1.2 (Chetty et al., 2011). To make hours worked one third of total hours, the leisure preference parameter is adjusted to $\chi = 1.73$. In this exercise, we find that the optimal capital tax rate is 8.1%, which is slightly higher than our previous result using a log-utility function. Hence, our main result of a positive optimal capital tax is robust to an iso-elastic labor supply.

[Figure 10 here]

Moreover, Figure 10 shows that the optimal capital tax rate is increasing with the Frisch elasticity of labor supply, which is consistent with the result of Aghion et al. (2013). We also find that the optimal capital tax rate becomes negative when the Frisch elasticity is very small ($1/\theta < 0.27$). The intuition is quite clear. When labor supply is more elastic, taxing labor income brings about a larger distortion. Therefore, the government tends to shift from the labor income tax to the capital income tax.

5.4 Alternative labor-leisure allocations

In our baseline calibration, we follow the standard literature in targeting work time equal to one third of total hours, which implies a labor-leisure allocation of 1:2. In this extension, we consider different labor-leisure allocations to explore their implications for the optimal capital income tax. In doing so, we recalibrate the parameter reflecting the leisure preference χ. We consider two alternative cases in addition to the benchmark. In the case of a high work time, we match work time equal to 0.5 of total hours (a labor-leisure ratio of 1:1), which gives us $\chi = 0.797$. In the case of a low work time, we match work time equal to 0.25 of total hours (a labor-leisure ratio of 1:3), which gives us $\chi = 3.403$.

[Table 2 here]

Table 2 reports the corresponding optimal capital tax rates. As shown, the optimal capital tax rate is lower in the case where households work more. This is because when the labor supply is high, it is less elastic, and taxing it causes less distortion. Therefore, the government tends to shift from the capital income tax to the labor income tax. Moreover, since the early 1970s, the European economies significantly worked less than the American (Maoz, 2010). According to Rogerson and Wallenius (2009), the aggregate working hours in continental Europe only amounts to about 70% of that in the US. Given that our benchmark targets the US economy, the amount of working hours in EU countries is close to the case of low work time. As such, Table 2 primarily implies that the optimal capital tax rate should be higher in EU countries. Undoubtedly, a comprehensive cross-country comparison of the optimal capital tax necessitates a more careful calibration of other important parameters. Since our main focus is on the linkage between R&D externalities and optimal capital taxation, we leave this issue for a future study.

6 Conclusion

In this paper, we have examined whether the Chamley-Judd result of zero optimal capital taxation is valid in a non-scale innovation-based growth model. By calibrating our model to the US economy, our result shows that the optimal capital income tax is positive, at a rate of around 6.6%. We examine how the optimal capital tax rate responds to various R&D externalities. The optimal capital tax rate is higher when (i) the "stepping on toes effect" is smaller, (ii) the "standing on shoulders effect" is stronger, or (iii) the extent of creative destruction is smaller. We also find that the optimal capital tax is sensitive to the parameter that determines the monopolistic markup. An inverted-U relationship is found between these two variables.

Some extensions for future study are worth noting. First, since R&D investment usually has liquidity problems (Lach, 2002), it would be relevant to introduce a credit constraint on R&D investment into our model. Moreover, it would be interesting to examine the optimal capital tax in an endogenous growth model where both innovation and capital accumulation are the driving forces of economic growth (see, e.g., Iwaisako and Futagami, 2013; Chu et al., 2019). These directions will generate new insights into the debate on the Chamley-Judd result.

References

Acemoglu, D. (1998) Why do new technologies complement skills? Directed technical change and wage inequality. *Quarterly Journal of Economics* 113, 1055-1089.

Aghion, P. and P. Howitt (1992) A model of growth through creative destruction. *Econometrica* 60, 323-351.

Aghion, P., U. Akcigit and J. Fernndez-Villaverde (2013) Optimal capital versus labor taxation with innovation-led growth. *NBER Working Paper* 19086.

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

Andolfatto D., S. Hendry and K. Moran (2008). Are inflation expectations rational?. *Journal of Monetary Economics* 55, 406-422.

Ang, J. B. and J. B. Madsen (2015) What drives ideas production across the world?. *Macroeconomic Dynamics* 19, 79-115.

Arnold, J. M., B. Brys, C. Heady, A. Johansson, C. Schwellnus and L. Vartia (2011) Tax policy for economic recovery and growth. *Economic Journal* 121, F59-F80.

Atkeson, A., V. V. Chari and P. J. Kehoe (1999) Taxing capital income: A bad idea. *Federal Reserve Bank of Minneapolis Quarterly Review* 23, 3-17.

Bloom, N., M. Schankerman and J. V. Reenen (2013) Identifying technology spillovers and product market rivalry. *Econometrica* 81, 1347-1393.

Carey, D. and H. Tchilingurian (2000) Average effective tax rates on capital, labor and consumption. *OECD Working Paper* 258.

Chamley, C. (1986) Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica* 54, 607-622.

Chamley, C. (2001) Capital income taxation, wealth redistribution and borrowing constraints. *Journal of Public Economics* 79, 55-69.

Chang, C. W. and C. C. Lai (2020) A note on optimal capital taxation with preference externalities. *Macroeconomic Dynamics* 24, 729-746.

Chari, V. V., J. P. Nicolini and P. Teles (2020) Optimal capital taxation revisited. *Journal of Monetary Economics* 116, 147-165.

Chari, V. V., L. J. Christiano and P. J. Kehoe (1994) Optimal fiscal policy in a business cycle model. *Journal of Political Economy* 102, 67-622.

Chari, V. V. and P. J. Kehoe (1990) Sustainable plans. *Journal of Political Economy* 98, 783-802.

Chari, V. V. and P. J. Kehoe (1999) Optimal fiscal and monetary policy. In: Taylor, J. B. and M. Woodford, (Eds.), *Handbook of Macroeconomics* 1C. Amsterdam: North-Holland, 1671-1745.

Chen, B. L., H. J. Chen and P. Wang (2021) Long-run tax incidence in a human-capital-based endogenous growth model with labor-market frictions. *Macroeconomic Dynamics* 25, 337-378.

Chen, B. L. and C. H. Lu (2013) Optimal factor tax incidence in two-sector human capital-based models. *Journal of Public Economics* 97, 75-94.

Chen, P., A. Chu, H. Chu and C. C. Lai (2019) Optimal capital taxation in an economy with innovation-driven growth. *MPRA Working Paper* 101228.

Chen, S. H. and J. T. Guo (2013) Progressive taxation and macroeconomic (In) stability with productive government spending. *Journal of Economic Dynamics and Control* 37, 951-963.

Chetty, R., A. Guren, D. Manoli and A. Weber (2011) Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins. *American Economic Review* 101, 471-475.

Chu, A., C. C. Lai and C. H. Liao (2019). A tale of two growth engines: Interactive effects of monetary policy and intellectual property rights. *Journal of Money, Credit and Banking*

51, 2029-2052.

Chu, A. and G. Cozzi (2014) R&D and economic growth in a cash-in-advance economy. *International Economic Review* 55, 507-524.

Chu, A. and G. Cozzi (2018) Effects of patents versus R&D subsidies on income inequality. *Review of Economic Dynamics* 29, 68-84.

Comin, D. (2004) R&D: A small contribution to productivity growth. *Journal of Economic Growth* 9, 391-421.

Conesa, J. C., S. Kitao and D. Krueger (2009) Taxing capital? Not a bad idea after all!. *American Economic Review* 99, 25-38.

Coto-Martinez, J., F. Sanchez-Losada and C. Garriga (2007) Optimal taxation with imperfect competition and aggregate returns to specialization. *Journal of the European Economic Association* 5, 1269-1299.

Cozzi, G. (2004) Rat race, redistribution, and growth. *Review of Economic Dynamics* 7, 900-915.

Domeij, D. (2005) Optimal capital taxation and labor market search. *Review of Economic Dynamics* 8, 623-650.

Eicher, T. S. and S. J. Turnovsky (2001) Transitional dynamics in a two-sector non-scale growth model. *Journal of Economic Dynamics and Control* 25, 85-113.

Einiö, E. (2014) R&D subsidies and company performance: Evidence from geographic variation in government funding based on the ERDF population-density rule. *Review of Economics and Statistics* 96, 710-728.

Erosa, A. and M. Gervais (2002) Optimal taxation in life-cycle economies. *Journal of Economic Theory* 105, 338-369.

Farhi, E. and I. Werning (2010) Progressive estate taxation. *Quarterly Journal of Economics* 125, 635-673.

Farhi, E. and I. Werning (2013) Estate taxation with altruism heterogeneity. *American*

Economic Review 103, 489-495.

Gali, J. (1994) Government size and macroeconomic stability. *European Economic Review* 38, 117-132.

Guo, J. T. and K. J. Lansing (1998) Indeterminacy and stabilization policy. *Journal of Economic Theory* 82, 481-490.

Golosov, M., N. Kocherlakota and A. Tsyvinski (2003) Optimal indirect and capital taxation. *Review of Economic Studies* 70, 569-587.

Golosov, M. and A. Tsyvinski (2015) Policy implications of dynamic public finance. *Annual Reviews of Economics* 7, 147-171.

Golosov, M., A. Tsyvinski and I. Werning (2006) New dynamic public finance: A user's guide. *NBER Macroeconomic Annual*, 317-363.

Grossman, G. M. and E. Helpman (1991) Quality ladders in the theory of growth. *Review of Economic Studies* 58, 43-61.

Imrohoroglu, S. (1998) A quantitative analysis of capital income taxation. *International Economic Review* 39, 307-328.

Iwaisako, T. (2016) Effects of patent protection on optimal corporate income and consumption taxes in an R&D-based growth model. *Southern Economic Journal* 83, 590-608.

Iwaisako, T. and K. Futagami (2013) Patent protection, capital accumulation, and economic growth. *Economic Theory* 52, 631-668.

Jermann, U. and V. Quadrini (2012) Macroeconomic effects of financial shocks. *American Economic Review* 102, 238-271..

Jones, C. I. (1995) R&D-based models of economic growth. *Journal of Political Economy* 103, 759-784.

Jones, C. I. and J. C. Williams (1998) Measuring the social return to R&D. *Quarterly Journal of Economics* 113, 1119-1135.

Jones, C. I. and J. C. Williams (2000) Too much of a good thing? The economics of investment in R&D. *Journal of Economic Growth* 5, 65-85.

Jones, L., R. Manuelli and P. Rossi (1997) On the optimal taxation of capital income. *Journal of Economic Theory* 73, 93-117.

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

Judd, K. L. (1997) The optimal tax rate for capital is negative. *NBER Working Paper* 6004.

Judd, K. L. (2002) Capital-income taxation with imperfect competition. *American Economic Review* 92, 417-421.

Lach, S. (2002) Do R&D subsidies stimulate or displace private R&D? Evidence from Israel. *Journal of Industrial Economics* 50, 369-390.

Laitner, J. and D. Stolyarov (2004) Aggregate returns to scale and embodied technical change: Theory and measurement using stock market data. *Journal of Monetary Economics* 51, 191-233.

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

Long, X. and A. Pelloni (2017) Factor income taxation in a horizontal innovation model. *Journal of Public Economics* 154, 137-159.

Lu, C. C. and B. L. Chen (2015) Optimal capital taxation in a neoclassical growth model. *Journal of Public Economic Theory* 17, 257-269.

Lucas, R. E. (1990) Supply-side economics: An analytical review. *Oxford Economic Papers* 42, 293-316.

Maoz, Y. D. (2010) Labor hours in the United States and Europe: the role of different leisure preferences. *Macroeconomic Dynamics* 14, 231-241.

Mattesini, F. and L. Rossi. (2012). Monetary policy and automatic stabilizers: the role of progressive taxation. *Journal of Money, Credit and Banking* 44, 825-862.

Mirrlees, J. A. (1971) An exploration in the theory of optimum income taxation. *Review of*

Economic Studies 38, 175-208.

Neves, P. C. and T. N. Sequeira (2018) Spillovers in the production of knowledge: A meta-regression analysis. *Research Policy* 47, 750-767.

Petrucci, A. (2015) Optimal income taxation in models with endogenous fertility. *Journal of Macroeconomics* 43, 216-225.

Piketty, T. and E. Saez (2013) A theory of optimal inheritance taxation. *Econometrica* 81, 1851-1886.

Rogerson, R. and J. Wallenius (2009) Micro and macro elasticities in a life cycle model with taxes. *Journal of Economic Theory* 144, 2277-2292.

Romer, P. (1990) Endogenous technological progress. *Journal of Political Economy* 98, S71-S102.

Scrimgeour, D. (2015) Dynamic scoring in a Romer-style economy. *Southern Economic Journal* 81, 697-723.

Sequeira, T. N. and P. C. Neves (2020) Stepping on toes in the production of knowledge: A meta-regression analysis. *Applied Economics* 52, 260-274.

Stokey, N. L. (1991) Credible public policy. *Journal of Economic Dynamics and Control* 15, 627-656.

Straub, L. and I. Werning (2020) Positive long-run capital taxation: Chamley-Judd revisited. *American Economic Review* 110, 86-119.

Yang, Y. (2018) On the optimality of IPR protection with blocking patents. *Review of Economic Dynamics* 27, 205-230.

Zeng, J. and J. Zhang (2002) Long-run growth effects of taxation in a non-scale growth model with innovation. *Economics Letters* 75, 391-403.

Appendix A. Deriving the steady-state growth rate

To solve for the steady-state growth rate of the economy, from (13) and (14) we have:

$$\frac{\dot{A}_t}{A_t} = \frac{1}{1+\psi}\frac{L_{A,t}^{\lambda}}{A_t^{1-\phi}}. \tag{A1}$$

where $g_{A,t} = \dot{A}_t/A_t$. Let g_Z denote $g_{Z,t} = \frac{\dot{Z}}{Z}$ the growth rate of any generic variable Z, and drop the time subscript when referring to any variables in the steady state. The steady-state growth rate of varieties is given by:

$$g_A = \frac{1}{1+\psi}\frac{L_A^{\lambda}}{A^{1-\phi}}. \tag{A2}$$

Moreover, the R&D labor share is $s_t = L_{A,t}/(N_t l_t)$. In so doing, (A2) can alternatively be expressed as:

$$g_A = \frac{1}{1+\psi}\frac{(sNl)^{\lambda}}{A^{1-\phi}}. \tag{A3}$$

By taking logarithms of (A3) and differentiating the resulting equation with respect to time, we have the following steady-state expression:

$$g_A = \frac{\lambda}{1-\phi}n. \tag{A4}$$

Equipped with the symmetric feature $x(i) = x$, the equilibrium condition for the capital market $K = Av$, and the production in the intermediate-good sector $x = v$, the aggregate production function can be rewritten as:

$$Y_t = A_t^{\frac{1}{\rho}-\alpha} L_t^{\alpha} K_t^{1-\alpha}. \tag{A5}$$

Taking logarithms of (A5) and differentiating the resulting equation with respect to time, we can infer the following result:

$$g_Y = \frac{(\frac{1}{\rho} - \alpha)}{1-\alpha} g_A + n. \tag{A6}$$

Inserting (A4) into (A6) yields:

$$g_Y = \sigma n, \tag{A7}$$

where $\sigma \equiv 1 + \frac{(\frac{1}{\rho}-\alpha)}{1-\alpha}\frac{\lambda}{1-\phi}$ is a composite parameter.

We now turn to solve the steady-state R&D labor share. In the long run, substituting $\dot{A}_t = g_A A_t$ and differentiating the resulting equation with respect to time gives rise to:

$$\dot{P}_A / P_A = g_Y - g_A \tag{A8}$$

From (12), (15), and (17), in the steady state we have:

$$\begin{aligned}
\pi_x &= \frac{\eta - 1}{\eta}\alpha\frac{Y}{A} && \text{(A9)}\\
P_A &= \frac{s}{1-s}\frac{(1-\alpha)Y/A}{(1+\psi)g_A} && \text{(A10)}\\
r &= \frac{\pi_x}{P_A} + \frac{\dot{P}_A}{P_A} - \psi g_A && \text{(A11)}
\end{aligned}$$

Substituting (A8), (A9), and (A10) into (A11) yields the result:

$$r = \frac{\frac{\eta-1}{\eta}\alpha Y/A}{\frac{s}{1-s}\frac{(1-\alpha)Y/A}{(1+\psi)g_A}} + g_Y - (1+\psi)g_A \tag{A12}$$

Based on (A12), we have the stationary R&D labor share s as follows:

$$s = \frac{\frac{\eta-1}{\eta}\frac{\alpha}{1-\alpha}(1+\psi)g_A}{r - g_Y + (1 + \frac{\eta-1}{\eta}\frac{\alpha}{1-\alpha})(1+\psi)g_A} \tag{A13}$$

Appendix B. Transition dynamics

This appendix solves the dynamic system of the model under *tax shifting from labor income taxes to capital income taxes*. The set of equations under the model is expressed by:

$$
\begin{aligned}
\frac{1}{c_t} &= q_t, && \text{(B1)}\\
\chi &= q_t(1-\tau_{L,t})w_t(1-l_t), && \text{(B2)}\\
r_t &= (1-\tau_K)r_{K,t}-\delta, && \text{(B3)}\\
\frac{\dot{c}_t}{c_t} &= r_t-n-\beta, && \text{(B4)}\\
w_t &= (1-\alpha)\frac{Y_t}{L_{Y,t}}, && \text{(B5)}\\
\eta r_{K,t} &= \alpha A_t^{\frac{1}{\rho}-1}L_{Y,t}^{1-\alpha}x_t^{\alpha-1}, && \text{(B6)}\\
r_{K,t}K_t &= \frac{\alpha}{\eta}Y_t, && \text{(B7)}\\
\pi_{x,t} &= \frac{\eta-1}{\eta}\alpha\frac{Y_t}{A_t}, && \text{(B8)}\\
r_tP_{A,t} &= \pi_{x,t}+\dot{P}_{A,t}-\psi\frac{\dot{A}_t}{A_t}P_{A,t}, && \text{(B9)}\\
G_t &= \zeta Y_t, && \text{(B10)}\\
G_t &= N_t(\tau_K r_{K,t}k_t+\tau_{L,t}w_tl_t), && \text{(B11)}\\
Y_t &= A_t^{1/\rho-\alpha}L_{Y,t}^{1-\alpha}K_t^{\alpha}, && \text{(B12)}\\
\dot{K}_t &= Y_t-C_t-G_t-\delta K_t, && \text{(B13)}\\
\frac{\dot{A}_t}{A_t} &= \frac{1}{1+\psi}\frac{L_{A,t}^{\lambda}}{A_t^{1-\phi}}, && \text{(B14)}\\
P_{A,t} &= \frac{s_t}{1-s_t}\frac{(1-\alpha)Y_t}{(1+\psi)\dot{A}_t}, && \text{(B15)}\\
N_tl_t &= L_{Y,t}+L_{A,t}. && \text{(B16)}
\end{aligned}
$$

The above 16 equations determine 16 unknowns $\{c_t,\ l_t,\ A_t,\ K_t,\ L_{Y,t},\ x_t,\ r_{K,t},\ \pi_{x,t}\ r_t,\ G_t,\ \tau_{L,t},\ Y_t,\ q_t,\ L_{A,t},\ P_{A,t},\ w_t\}$, where q_t is the Hamiltonian multiplier, $C_t = N_t c_t$, $K_t \equiv N_t k_t = A_t x_t$, and $s_t = L_{A,t}/N_t l_t$. Based on $K_t = N_t k_t = A_t x_t$, and (B1), (B2), (B5), and (B12), we can obtain:

$$\chi = \frac{1}{c_t}(1-\tau_{L,t})(1-\alpha)\frac{Y_t}{L_{Y,t}}(1-l_t). \tag{B17a}$$

From (B5), (B7), and (B11), we have:

$$\tau_{L,t} = (1-s_t)\frac{\zeta - \frac{\alpha}{\eta}\tau_K}{1-\alpha}. \tag{B17b}$$

Moreover, to solve the balanced growth rate, we define the following transformed variables:

$$\hat{k}_t \equiv \frac{K_t}{N_t^\sigma}, \quad \hat{c}_t \equiv \frac{C_t}{N_t^\sigma}, \quad \hat{y}_t \equiv \frac{Y_t}{N_t^\sigma}, \quad \hat{a}_t \equiv \frac{A_t}{N_t^{\lambda/(1-\phi)}}, \quad s_t \equiv L_{A,t}/N_t l_t. \tag{B18}$$

Based on (B16), (B15), (B17a), and the above definitions, we can obtain:

$$\frac{\chi}{(1-l_t)} = \frac{1}{\hat{c}_t}[1-(1-s_t)\frac{\zeta - \frac{\alpha}{\eta}\tau_K}{1-\alpha}](1-\alpha)\hat{a}_t^{1/\rho-\alpha}(\hat{k}_t)^\alpha[1-s_t)l_t]^{-\alpha}. \tag{B19a}$$

From (B19a), we can infer the following expression:

$$l_t = l_t(\hat{k}_t, \hat{a}_t, \hat{c}_t, s_t; \tau_K), \tag{B19b}$$

where

$$\frac{\partial l_t}{\partial \hat{k}_t} = \frac{\alpha}{\hat{k}_t(\frac{l_t}{1-l_t}+\alpha)} l_t, \tag{B20a}$$

$$\frac{\partial l_t}{\partial \hat{a}_t} = \frac{(1/\rho-\alpha)}{\hat{a}_t(\frac{l_t}{1-l_t}+\alpha)} l_t, \tag{B20b}$$

$$\frac{\partial l_t}{\partial \hat{c}_t} = -\frac{l_t}{\hat{c}_t(\frac{l_t}{1-l_t}+\alpha)}, \tag{B20c}$$

$$\frac{\partial l_t}{\partial s_t} = \frac{\frac{\tau_{L,t}}{(1-\tau_{L,t})}+\alpha}{(1-s_t)(\frac{l_t}{1-l_t}+\alpha)} l_t, \tag{B20d}$$

$$\frac{\partial l_t}{\partial \tau_K} = \frac{(1-s_t)\frac{\alpha}{\eta(1-\alpha)}}{(1-\tau_{L,t})(\frac{l_t}{1-l_t}+\alpha)} l_t. \tag{B20e}$$

Based on (B3), (B4), (B7), (B12), (B18), and $C_t = N_t c_t$, we have:

$$g_{\hat{c},t} \equiv \frac{d\hat{c}_t/dt}{\hat{c}_t} = (1-\tau_K)\frac{\alpha}{\eta}(\hat{a}_t)^{1/\rho-\alpha}[\frac{(1-s_t)l_t(\hat{k}_t,\hat{a}_t,\hat{c}_t,s_t;\tau_K)}{\hat{k}_t}]^{1-\alpha} - \delta - \beta - g_Y. \tag{B21}$$

From (B10), (B12), (B13), and (B18), we can directly infer:

$$g_{\hat{k},t} \equiv \frac{d\hat{k}_t/dt}{\hat{k}_t} = (1-\zeta)(\hat{a}_t)^{1/\rho-\alpha}[\frac{(1-s_t)l_t(\hat{k}_t,\hat{a}_t,\hat{c}_t,s_t;\tau_K)}{\hat{k}_t}]^{1-\alpha} - \frac{\hat{c}_t}{\hat{k}_t} - \delta - g_Y. \tag{B22}$$

According to (B14) and (B18), we can further obtain:

$$g_{\hat{a},t} \equiv \frac{d\hat{a}_t/dt}{\hat{a}_t} = \frac{1}{1+\psi}\frac{[s_t l_t(\hat{k}_t,\hat{a}_t,\hat{c}_t,s_t;\tau_K)]^{\lambda}}{\hat{a}_t^{1-\phi}} - g_A. \tag{B23}$$

In what follows, to simplify the notation we suppress those arguments of the labor supply function. From (B18), taking logarithms of (B19a) and (B12) and differentiating the resulting equations with respect to time, we have:

$$g_{\hat{y},t} = (1/\rho - \alpha)g_{\hat{a},t} + \alpha g_{\hat{k},t} + (1-\alpha)(\dot{l}_t/l_t - \frac{\dot{s}_t}{1-s_t}), \tag{B24}$$

$$\dot{l}_t/l_t = \{(1/\rho - \alpha)g_{\hat{a},t} + \alpha g_{\hat{k},t} - g_{\hat{c},t} - [\alpha + \tau_{L,t}/(1-\tau_{L,t})]\}/[\alpha + l_t/(1-l_t)]. \tag{B25}$$

Taking logarithms of (B15) differentiating the resulting equation with respect to time, we obtain:

$$\frac{\dot{P}_{A,t}}{P_{A,t}} = (1/\rho - \alpha - \phi)g_{\hat{a},t} + \alpha g_{\hat{k},t} + (1 - \lambda + \alpha\frac{s_t}{1-s_t})\frac{\dot{s}_t}{s_t} + (1-\lambda-\alpha)\frac{\dot{l}_t}{l_t} + g_Y - g_A. \tag{B26}$$

Combining (B9), (B15), (B18), (B21), (B24), (B25), and (B26) together, we obtain:

$$\begin{aligned}\frac{ds_t/dt}{s_t} \quad = \quad & \{\beta - [\frac{(\eta-1)\alpha(1+\psi)(1-s_t)}{(1-\alpha)\eta s_t} - \psi](g_A + g_{\hat{a},t}) + \phi g_{\hat{a},t} + g_A - [1 + \frac{1-\lambda-\alpha}{\alpha + l_t/(1-l_t)}] \\ & \times[(1/\rho - \alpha)g_{\hat{a},t} + \alpha g_{\hat{k},t} - g_{\hat{c},t}]\}/\{1 - \lambda + \alpha\frac{s_t}{1-s_t} + \frac{1-\lambda-\alpha}{\alpha + l_t/(1-l_t)}(\alpha + \frac{\tau_{L,t}}{1-\tau_{L,t}})\frac{s_t}{1-s_t}\}.\end{aligned} \tag{B27}$$

Note that $r_t - g_Y - g_{\hat{c},t} = \beta$. As a result, in the steady state we have $r - g_Y = \beta$.

Inserting (B18) into (B17b) yields:

$$\tau_{L,t} = (1-s_t)\frac{\zeta - \frac{\alpha}{\eta}\tau_K}{1-\alpha}. \tag{B28}$$

Based on (B21), (B22), (B23),(B27), and (B28), the dynamic system can be expressed as:

$$\frac{d\hat{k}_t/dt}{\hat{k}_t} = (1-\zeta)(\hat{a}_t)^{1/\rho-\alpha}[\frac{(1-s_t)l_t}{\hat{k}_t}]^{1-\alpha} - \frac{\hat{c}_t}{\hat{k}_t} - \delta - g_Y, \tag{B29a}$$

$$\frac{d\hat{a}_t/dt}{\hat{a}_t} = \frac{1}{1+\psi}\frac{(s_t l_t)^{\lambda}}{\hat{a}_t^{1-\phi}} - g_A, \tag{B29b}$$

$$\frac{d\hat{c}_t/dt}{\hat{c}_t} = (1-\tau_K)\frac{\alpha}{\eta}(\hat{a}_t)^{1/\rho-\alpha}[\frac{(1-s_t)l_t}{\hat{k}_t}]^{1-\alpha} - \delta - \beta - g_Y, \tag{B29c}$$

$$\begin{aligned}\frac{ds_t/dt}{s_t} = {} & \{\beta - [\frac{(\eta-1)\alpha(1+\psi)(1-s_t)}{(1-\alpha)\eta s_t} - \psi](g_A + g_{\hat{a},t}) + \phi g_{\hat{a},t} + g_A - [1 + \frac{1-\lambda-\alpha}{\alpha + l_t/(1-l_t)}] \\ & \times[(1/\rho-\alpha)g_{\hat{a},t} + \alpha g_{\hat{k},t} - g_{\hat{c},t}]\}/\{1-\lambda+\alpha\frac{s_t}{1-s_t} + \frac{1-\lambda-\alpha}{\alpha+l_t/(1-l_t)}(\alpha + \frac{\tau_{L,t}}{1-\tau_{L,t}})\frac{s_t}{1-s_t}\}.\end{aligned} \tag{B29d}$$

Linearizing (B29a), (B29b), (B29c), and (B29d) around the steady-state equilibrium yields:

$$\begin{pmatrix} d\hat{k}_t/dt \\ d\hat{a}_t/dt \\ d\hat{c}_t/dt \\ ds_t/dt \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{pmatrix} \begin{pmatrix} \hat{k}_t - \hat{k} \\ \hat{a}_t - \hat{a} \\ \hat{c}_t - \hat{c} \\ s_t - s \end{pmatrix} + \begin{pmatrix} b_{15} \\ b_{25} \\ b_{35} \\ b_{45} \end{pmatrix} d\tau_K, \tag{B30}$$

where

$b_{11} = \frac{\partial(d\hat{k}_t/dt)}{\partial\hat{k}_t}$, $b_{12} = \frac{\partial(d\hat{k}_t/dt)}{\partial\hat{a}_t}$, $b_{13} = \frac{\partial(d\hat{k}_t/dt)}{\partial\hat{c}_t}$, $b_{14} = \frac{\partial(d\hat{k}_t/dt)}{\partial s_t}$, $b_{15} = \frac{\partial(d\hat{k}_t/dt)}{\partial\tau_K}$,
$b_{21} = \frac{\partial(d\hat{a}_t/dt)}{\partial\hat{k}_t}$, $b_{22} = \frac{\partial(d\hat{a}_t/dt)}{\partial\hat{a}_t}$, $b_{23} = \frac{\partial(d\hat{a}_t/dt)}{\partial\hat{c}_t}$, $b_{24} = \frac{\partial(d\hat{a}_t/dt)}{\partial s_t}$, $b_{25} = \frac{\partial(d\hat{a}_t/dt)}{\partial\tau_K}$,
$b_{31} = \frac{\partial(d\hat{c}_t/dt)}{\partial\hat{k}_t}$, $b_{32} = \frac{\partial(d\hat{c}_t/dt)}{\partial\hat{a}_t}$, $b_{33} = \frac{\partial(d\hat{c}_t/dt)}{\partial\hat{c}_t}$, $b_{34} = \frac{\partial(d\hat{c}_t/dt)}{\partial s_t}$, $b_{35} = \frac{\partial(d\hat{c}_t/dt)}{\partial\tau_K}$,
$b_{41} = \frac{\partial(ds_t/dt)}{\partial\hat{k}_t}$, $b_{42} = \frac{\partial(ds_t/dt)}{\partial\hat{a}_t}$, $b_{43} = \frac{\partial(ds_t/dt)}{\partial\hat{c}_t}$, $b_{44} = \frac{\partial(ds_t/dt)}{\partial s_t}$, $b_{45} = \frac{\partial(ds_t/dt)}{\partial\tau_K}$.

Due to the complicated calculations, we do not list the analytical results for b_{ij}, where $i \in \{1, 2, 3, 4, 5\}$ and $j \in \{1, 2, 3, 4, 5\}$.

Let ℓ_1, ℓ_2, ℓ_3, and ℓ_4 be the four characteristic roots of the dynamic system. Due to the complexity involved in calculating the four characteristic roots, we do not try to prove the saddle-point stability analytically. Instead, via a numerical simulation, we show that the dynamic system has two positive and two negative characteristic roots. For expository convenience, in what follows let ℓ_1 and ℓ_2 be the negative root, and ℓ_3 and ℓ_4 be the positive roots. The general solution is given by:

$$\begin{pmatrix} \hat{k}_t \\ \hat{a}_t \\ \hat{c}_t \\ s_t \end{pmatrix} = \begin{pmatrix} \hat{k}(\tau_K) \\ \hat{a}(\tau_K) \\ \hat{c}(\tau_K) \\ s(\tau_K) \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 & 1 \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ h_{41} & h_{42} & h_{43} & h_{44} \end{pmatrix} \begin{pmatrix} D_1 e^{\ell_1 t} \\ D_2 e^{\ell_2 t} \\ D_3 e^{\ell_3 t} \\ D_4 e^{\ell_4 t} \end{pmatrix}. \tag{B31a}$$

where D_1, D_2, D_3, and D_4 are undetermined coefficients and

$$\triangle_j = \begin{vmatrix} b_{12} & b_{13} & b_{14} \\ b_{22}-\ell_j & b_{23} & b_{24} \\ b_{32} & b_{33}-\ell_j & b_{34} \end{vmatrix} \; ; j \in \{1,\ 2,\ 3,\ 4\}, \tag{B31b}$$

$$h_{2j} = \begin{vmatrix} \ell_j - b_{11} & b_{13} & b_{14} \\ -b_{21} & b_{23} & b_{24} \\ -b_{31} & b_{33}-\ell_j & b_{34} \end{vmatrix} / \triangle_j \;\; ; j \in \{1,\ 2,\ 3,\ 4\}, \tag{B31c}$$

$$h_{3j} = \begin{vmatrix} b_{12} & -b_{11} & b_{14} \\ b_{22}-\ell_j & -b_{21} & b_{24} \\ b_{32} & -b_{31} & b_{34} \end{vmatrix} / \triangle_j \;\; ; j \in \{1,\ 2,\ 3,\ 4\}, \tag{B31d}$$

$$h_{4j} = \begin{vmatrix} b_{12} & b_{13} & \ell_j - b_{11} \\ b_{22}-\ell_j & b_{23} & -b_{21} \\ b_{32} & b_{33}-\ell_j & -b_{31} \end{vmatrix} / \triangle_j \;\; ; j \in \{1,\ 2,\ 3,\ 4\}. \tag{B31e}$$

The government changes the capital tax rate τ_K from τ_{K0} to τ_{K1} at t=0. Based on (B31a)-(B31e), we employ the following equations to describe the dynamic adjustment of $\hat{k}_t$, $\hat{a}_t$, $\hat{c}_t$ and s_t:

$$\hat{k}_t = \begin{cases} \hat{k}(\tau_{K0}); & t = 0^- \\ \hat{k}(\tau_{K1}) + D_1 e^{\ell_1 t} + D_2 e^{\ell_2 t} + D_3 e^{\ell_3 t} + D_4 e^{\ell_4 t}; & t \geq 0^+ \end{cases} \tag{B32a}$$

$$\hat{a}_t = \begin{cases} \hat{a}(\tau_{K0}); & t = 0^- \\ \hat{a}(\tau_{K1}) + h_{21} D_1 e^{\ell_1 t} + h_{22} D_2 e^{\ell_2 t} + h_{23} D_3 e^{\ell_3 t} + h_{24} D_4 e^{\ell_4 t}; & t \geq 0^+ \end{cases} \tag{B32b}$$

$$\hat{c}_t = \begin{cases} \hat{c}(\tau_{K0}); & t = 0^- \\ \hat{c}(\tau_{K1}) + h_{31} D_1 e^{\ell_1 t} + h_{32} D_2 e^{\ell_2 t} + h_{33} D_3 e^{\ell_3 t} + h_{34} D_4 e^{\ell_4 t}; & t \geq 0^+ \end{cases} \tag{B32c}$$

$$s_t = \begin{cases} s(\tau_{K0}); & t = 0^- \\ s(\tau_{K1}) + h_{41} D_1 e^{\ell_1 t} + h_{42} D_2 e^{\ell_2 t} + h_{43} D_3 e^{\ell_3 t} + h_{44} D_4 e^{\ell_4 t}; & t \geq 0^+ \end{cases} \tag{B32d}$$

where 0^- and 0^+ denote the instant before and instant after the policy implementation, respectively. The values for D_1, D_2, D_3 and D_4 are determined by:

$$\hat{k}_{0^-} = \hat{k}_{0^+}, \tag{B33a}$$

$$\hat{a}_{0^-} = \hat{a}_{0^+}, \tag{B33b}$$

$$D_3 = D_4 = 0. \tag{B33c}$$

Equations (B33a) and (B33b) indicate that both $\hat{k}_t$ $(= \frac{K_t}{N_t^{\sigma}})$ and $\hat{a}_t$ $(= \frac{A_t}{N_t^{\lambda/(1-\phi)}})$ remain intact at the instant of policy implementation since K_t, A_t, and N_t are predetermined variables. Equation (B33c) is the stability condition which ensures that all $\hat{k}_t$, $\hat{a}_t$, $\hat{c}_t$ and s_t converge to their new steady-state equilibrium. By using (B33a) and (B33b), we can obtain:

$$D_1 = \frac{[\hat{k}(\tau_{K0}) - \hat{k}(\tau_{K1})]h_{22} - [\hat{a}(\tau_{K0}) - \hat{a}(\tau_{K1})]}{h_{22} - h_{21}}, \tag{B34a}$$

$$D_2 = \frac{[\hat{a}(\tau_{K0}) - \hat{a}(\tau_{K1})] - [\hat{k}(\tau_{K0}) - \hat{k}(\tau_{K1})]h_{21}}{h_{22} - h_{21}}. \tag{B34b}$$

Inserting (B33c), (B34a), and (B34b) into (B32a)-(B32d) yields:

$$\hat{k}_t = \begin{cases} \hat{k}(\tau_{K0}); & t = 0^- \\ \hat{k}(\tau_{K1}) + \frac{[\hat{k}(\tau_{K0})-\hat{k}(\tau_{K1})]h_{22}-[\hat{a}(\tau_{K0})-\hat{a}(\tau_{K1})]}{h_{22}-h_{21}} e^{\ell_1 t} & t \geq 0^+ \\ \qquad + \frac{[\hat{a}(\tau_{K0})-\hat{a}(\tau_{K1})]-[\hat{k}(\tau_{K0})-\hat{k}(\tau_{K1})]h_{21}}{h_{22}-h_{21}} e^{\ell_2 t}; & \end{cases}$$

$$\hat{a}_t = \begin{cases} \hat{a}(\tau_{K0}); & t = 0^- \\ \hat{a}(\tau_{K1}) + \frac{\{[\hat{k}(\tau_{K0})-\hat{k}(\tau_{K1})]h_{22}-[\hat{a}(\tau_{K0})-\hat{a}(\tau_{K1})]\}h_{21} e^{\ell_1 t}}{h_{22}-h_{21}} & t \geq 0^+ \\ \qquad + \frac{\{[\hat{a}(\tau_{K0})-\hat{a}(\tau_{K1})]-[\hat{k}(\tau_{K0})-\hat{k}(\tau_{K1})]h_{21}\}h_{22} e^{\ell_2 t}}{h_{22}-h_{21}}; & \end{cases}$$

$$\hat{c}_t = \begin{cases} \hat{c}(\tau_{K0}); & t = 0^- \\ \hat{c}(\tau_{K1}) + \frac{\{[\hat{k}(\tau_{K0})-\hat{k}(\tau_{K1})]h_{22}-[\hat{a}(\tau_{K0})-\hat{a}(\tau_{K1})]\}h_{31} e^{\ell_1 t}}{h_{22}-h_{21}} & t \geq 0^+ \\ \qquad + \frac{\{[\hat{a}(\tau_{K0})-\hat{a}(\tau_{K1})]-[\hat{k}(\tau_{K0})-\hat{k}(\tau_{K1})]h_{21}\}h_{32} e^{\ell_2 t}}{h_{22}-h_{21}}; & \end{cases}$$

$$s_t = \begin{cases} s(\tau_{K0}); & t = 0^- \\ s(\tau_{K1}) + \frac{\{[\hat{k}(\tau_{K0})-\hat{k}(\tau_{K1})]h_{22}-[\hat{a}(\tau_{K0})-\hat{a}(\tau_{K1})]\}h_{41} e^{\ell_1 t}}{h_{22}-h_{21}} & t \geq 0^+ \\ \qquad + \frac{\{[\hat{a}(\tau_{K0})-\hat{a}(\tau_{K1})]-[\hat{k}(\tau_{K0})-\hat{k}(\tau_{K1})]h_{21}\}h_{42} e^{\ell_2 t}}{h_{22}-h_{21}}; & \end{cases}$$

Appendix C. Proof of comparative statics

From (B29a)-(B29d), we have:

$$\frac{d\hat{k}_t/dt}{\hat{k}_t} = (1-\zeta)(\hat{a}_t)^{1/\rho-\alpha}(\frac{l_{Y,t}}{\hat{k}_t})^{1-\alpha} - \frac{\hat{c}_t}{\hat{k}_t} - \delta - g_Y, \tag{C1a}$$

$$\frac{d\hat{a}_t/dt}{\hat{a}_t} = \frac{1}{1+\psi}\frac{[l_t(\hat{k}_t, \hat{a}_t, \hat{c}_t, l_{Y,t}; \tau_K) - l_{Y,t}]^\lambda}{\hat{a}_t^{1-\phi}} - g_A, \tag{C1b}$$

$$\frac{d\hat{c}_t/dt}{\hat{c}_t} = (1-\tau_K)\frac{\alpha}{\eta}(\hat{a}_t)^{1/\rho-\alpha}(\frac{l_{Y,t}}{\hat{k}_t})^{1-\alpha} - \delta - \beta - g_Y, \tag{C1c}$$

$$\begin{aligned}\frac{ds_t/dt}{s_t} = {} & \{\beta - [\frac{(\eta-1)\alpha(1+\psi)(1-s_t)}{(1-\alpha)\eta s_t} - \psi](g_A + g_{\hat{a},t}) + \phi g_{\hat{a},t} + g_A - [1 + \frac{1-\lambda-\alpha}{\alpha + l_t/(1-l_t)}] \\ & \times[(1/\rho-\alpha)g_{\hat{a},t} + \alpha g_{\hat{k},t} - g_{\hat{c},t}]\}/\{1-\lambda+\alpha\frac{s_t}{1-s_t} + \frac{1-\lambda-\alpha}{\alpha+l_t/(1-l_t)}(\alpha + \frac{\tau_{L,t}}{1-\tau_{L,t}})\frac{s_t}{1-s_t}\}.\end{aligned} \tag{C1d}$$

In the steady state $\frac{d\hat{k}_t/dt}{\hat{k}_t} = \frac{d\hat{a}_t/dt}{\hat{a}_t} = \frac{d\hat{c}_t/dt}{\hat{c}_t} = \frac{ds_t/dt}{s_t} = 0$, we then have the following steady-state results:

$$\frac{\hat{c}}{\hat{k}} = (1-\zeta)(\hat{a})^{1/\rho-\alpha}[\frac{(1-s)l}{\hat{k}}]^{1-\alpha} - \delta - g_Y, \tag{C1e}$$

$$g_A = \frac{1}{1+\psi}\frac{(sl)^\lambda}{\hat{a}^{1-\phi}}, \tag{C1f}$$

$$\beta = (1-\tau_K)\frac{\alpha}{\eta}(\hat{a})^{1/\rho-\alpha}[\frac{(1-s)l}{\hat{k}}]^{1-\alpha} - \delta - g_Y, \tag{C1g}$$

$$0 = \beta - [\frac{(\eta-1)\alpha(1+\psi)(1-s)}{(1-\alpha)\eta s} - \psi]g_A + g_A. \tag{C1h}$$

Based on (C1h), we have:

$$s = \frac{\frac{\eta-1}{\eta}\frac{\alpha}{1-\alpha}(1+\psi)g_A}{\beta + \left(1 + \frac{\eta-1}{\eta}\frac{\alpha}{1-\alpha}\right)(1+\psi)g_A}. \tag{C2}$$

From (B3) and (C1g), we can obtain

$$r - g_Y = \beta > 0. \tag{C3}$$

Equation (C1g) can be rearranged as:

$$\hat{y}/\hat{k} = (\hat{a})^{1/\rho-\alpha}[\frac{(1-s)l}{\hat{k}}]^{1-\alpha} = \frac{\eta(\beta+\delta+g_Y)}{\alpha(1-\tau_K)}. \tag{C4a}$$

Substituting (C4a) into (C1e) gives rise to:

$$\frac{\hat{c}}{\hat{y}} = \{(1-\zeta)\frac{\eta(\beta+\delta+g_Y)}{\alpha(1-\tau_K)} - \delta - g_Y\}\frac{\hat{k}}{\hat{y}} = (1-\zeta) - (\delta+g_Y)\frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)}. \tag{C5a}$$

To ensure that the steady-state consumption-output ratio $\hat{c}/\hat{y}$ is positive, we impose the restriction $(1-\zeta) - (\delta+g_Y)\frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)} > 0$ for all values of the time preference rate β. As a consequence, $\lim_{\beta\to 0}\ \hat{c}/\hat{y}>0$ implies:

$$(1-\zeta) - \frac{\alpha(1-\tau_K)}{\eta} > 0. \tag{C5b}$$

From (C1f), we can derive:

$$\hat{a} = [\frac{1}{(1+\psi)g_A}]^{1/(1-\phi)}(sl)^{\lambda/(1-\phi)}. \tag{C6}$$

Based on (B28), we can infer the following expression:

$$\tau_L = (1-s)\frac{\zeta - \frac{\alpha}{\eta}\tau_K}{1-\alpha}, \tag{C7a}$$

where

$$\frac{\partial\tau_L}{\partial\tau_K} = -(1-s)\frac{\frac{\alpha}{\eta}}{1-\alpha} < 0. \tag{C7b}$$

Equipped with (B1), (B2), (B5), and $L_Y = N(1-s)l$, we can obtain:

$$\frac{l}{1-l}\chi = \frac{\hat{y}}{\hat{c}}\frac{(1-\tau_L)(1-\alpha)}{(1-s)}. \tag{C8}$$

Inserting (C5a) and (C7a) into (C8) yields:

$$l = \begin{cases} 1 - \frac{\chi}{\chi + \frac{1}{[(1-\zeta)-(\delta+g_Y)\frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)}]}\frac{(1-\tau_L)(1-\alpha)}{(1-s)}} & ; \quad \chi > 0 \\ 1 & ; \quad \chi = 0 \end{cases}, \tag{C9a}$$

where

$$\frac{\partial l}{\partial\tau_K} = \begin{cases} \frac{\alpha\beta(\frac{1-s}{1-\alpha})[1-\zeta+\frac{\eta-1}{\eta}\frac{\alpha(\delta+g_Y)}{\beta+(1+\psi)g_A}](1-l)l}{\eta(\beta+\delta+g_Y)(1-\tau_L)[1-\zeta-(\delta+g_Y)\frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)}]} > 0 & ; \quad \chi > 0 \\ 0 & ; \quad \chi = 0 \end{cases}. \tag{C9b}$$

Combining (C2), (C6), and (C9b) together, we can derive

$$\hat{a} = [\frac{1}{(1+\psi)g_A}]^{1/(1-\phi)}(sl)^{\lambda/(1-\phi)}, \tag{C10a}$$

where

$$\frac{\partial\hat{a}}{\partial\tau_K} = \frac{\lambda}{(1-\phi)}\hat{a}\frac{\partial l}{l\partial\tau_K} > 0. \tag{C10b}$$

Based on (C4a), (C9b), (B12), and (B18), we have:

$$\hat{y} = \hat{a}^{\frac{1/\rho-\alpha}{1-\alpha}}[\frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)}]^{\frac{\alpha}{1-\alpha}}(1-s)l, \tag{C11a}$$

where

$$\frac{\partial \hat{y}}{\partial \tau_K} = [\sigma \frac{\partial l}{l \partial \tau_K} - \frac{\alpha}{(1-\alpha)(1-\tau_K)}]\hat{y} \gtrless 0, \quad \sigma \equiv 1 + \frac{1/\rho - \alpha}{1-\alpha}\frac{\lambda}{1-\phi}. \tag{C11b}$$

According to (C4a), (C5a), and (C11b), we obtain:

$$\hat{k} = \frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)}\hat{y}, \tag{C12a}$$

$$\hat{c} = [(1-\zeta) - (\delta+g_Y)\frac{\alpha(1-\tau_K)}{\eta(\beta+\delta+g_Y)}]\hat{y}, \tag{C12b}$$

Inserting (C11a) into (C12a) and (C12b), we can derive the following comparative statics:

$$\frac{\partial \hat{k}}{\partial \tau_K} = \frac{\alpha(1-\tau_K)\hat{y}}{\eta(\beta+\delta+g_Y)}\{\sigma \frac{\partial l}{l \partial \tau_K} - \frac{1}{(1-\alpha)(1-\tau_K)}\} \gtrless 0, \tag{C12c}$$

$$\begin{aligned}\frac{\partial \hat{c}}{\partial \tau_K} = \{&\frac{\alpha(\delta+g_Y)}{\eta(\beta+\delta+g_Y)} + [(1-\zeta) \\ &- \frac{\alpha(1-\tau_K)(\delta+g_Y)}{\eta(\beta+\delta+g_Y)}][\sigma \frac{\partial l}{l \partial \tau_K} - \frac{\alpha}{(1-\alpha)(1-\tau_K)}]\}\hat{y} \gtrless 0.\end{aligned} \tag{C12d}$$

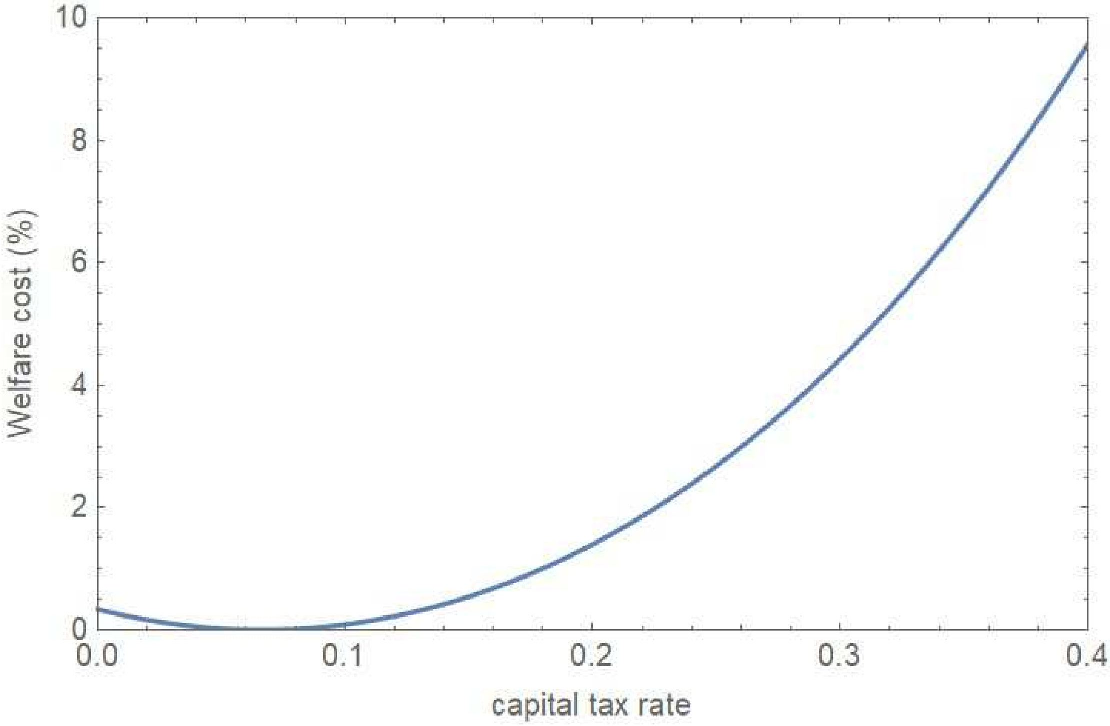

Figure 1: The optimal capital tax rate

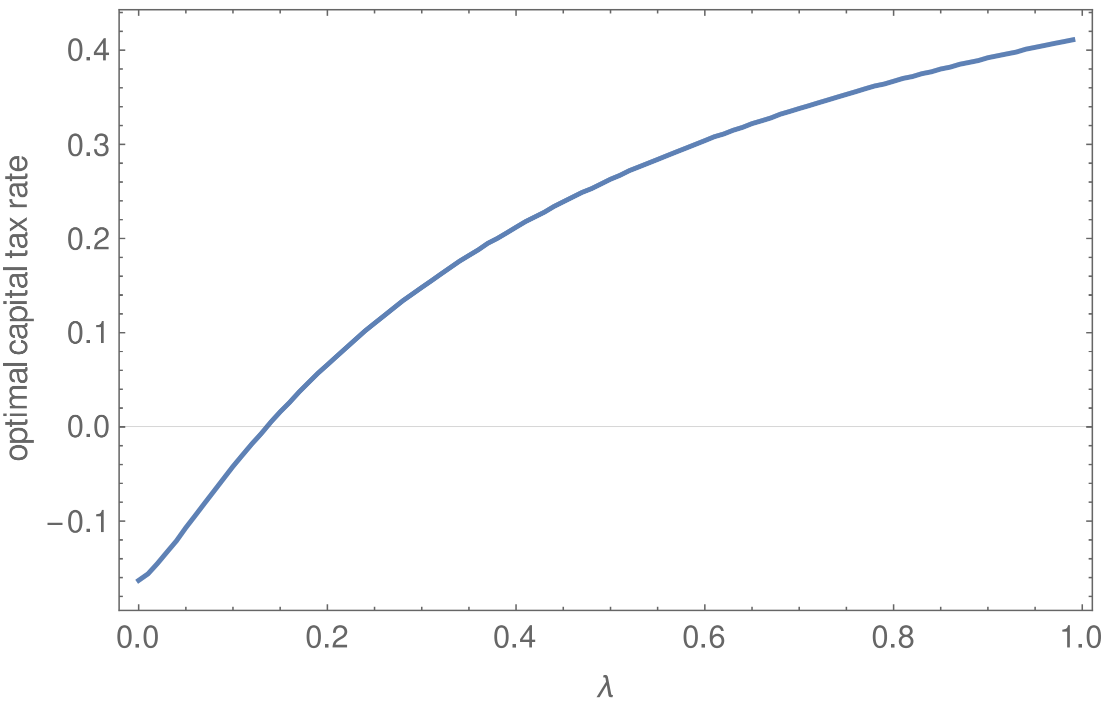

Figure 2: The optimal capital tax rate and the “stepping on toes effect”

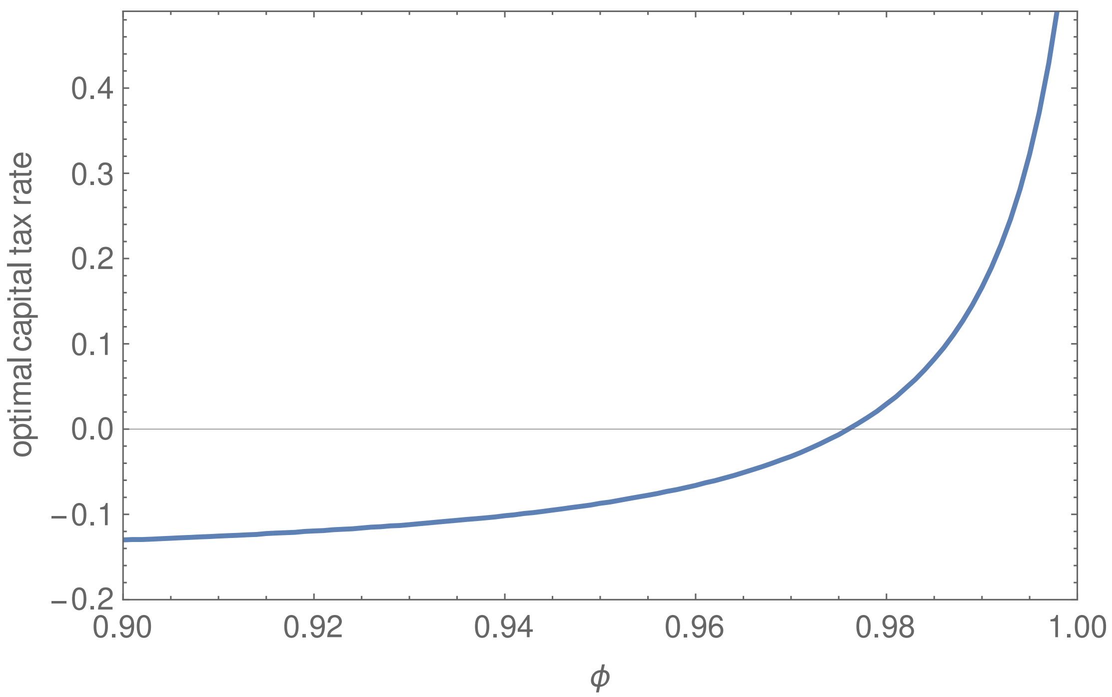

Figure 3: The optimal capital tax rate and the "standing on shoulders effect"

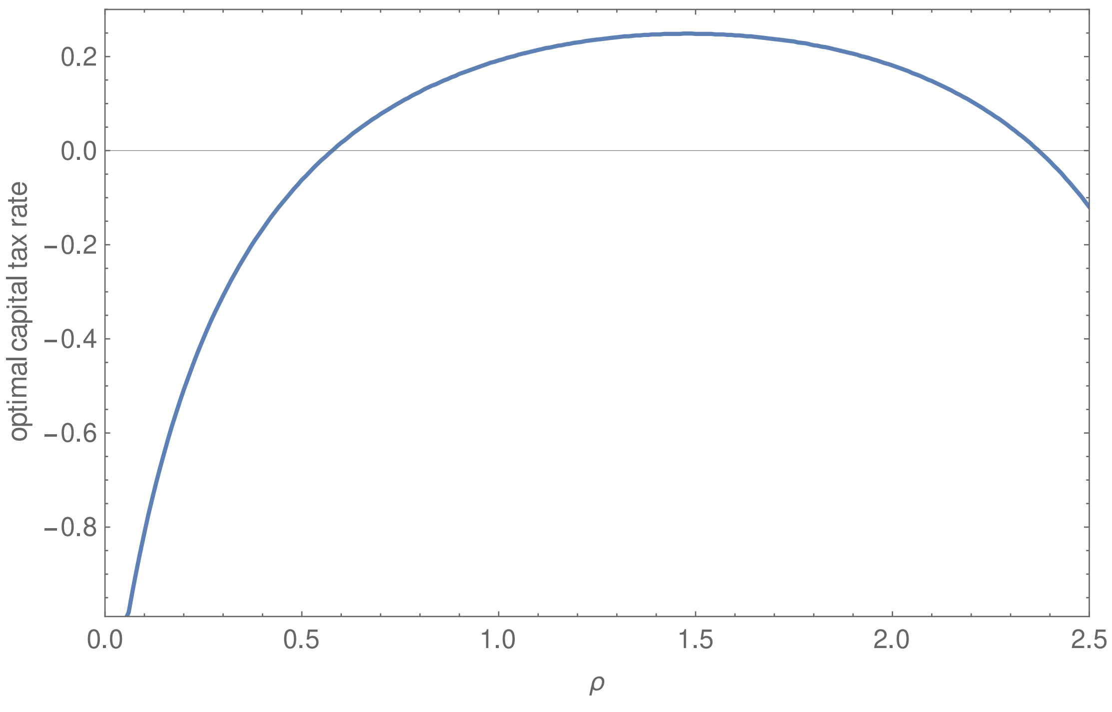

Figure 4: The optimal capital tax rate and the substitution parameter

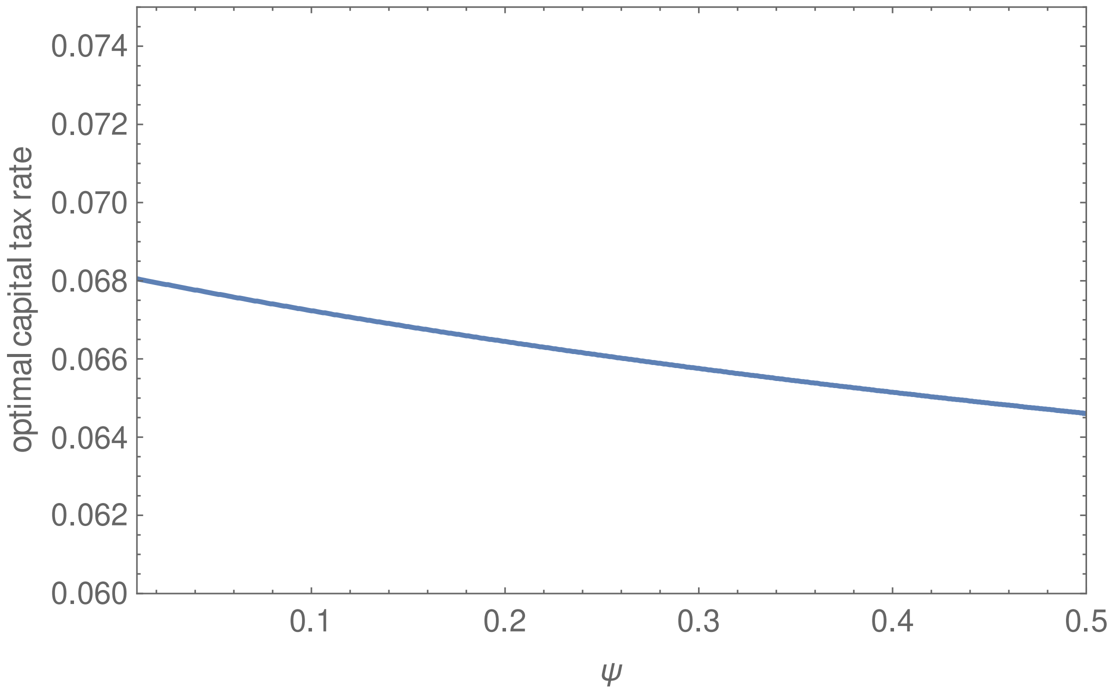

Figure 5: The optimal capital tax rate and creative destruction

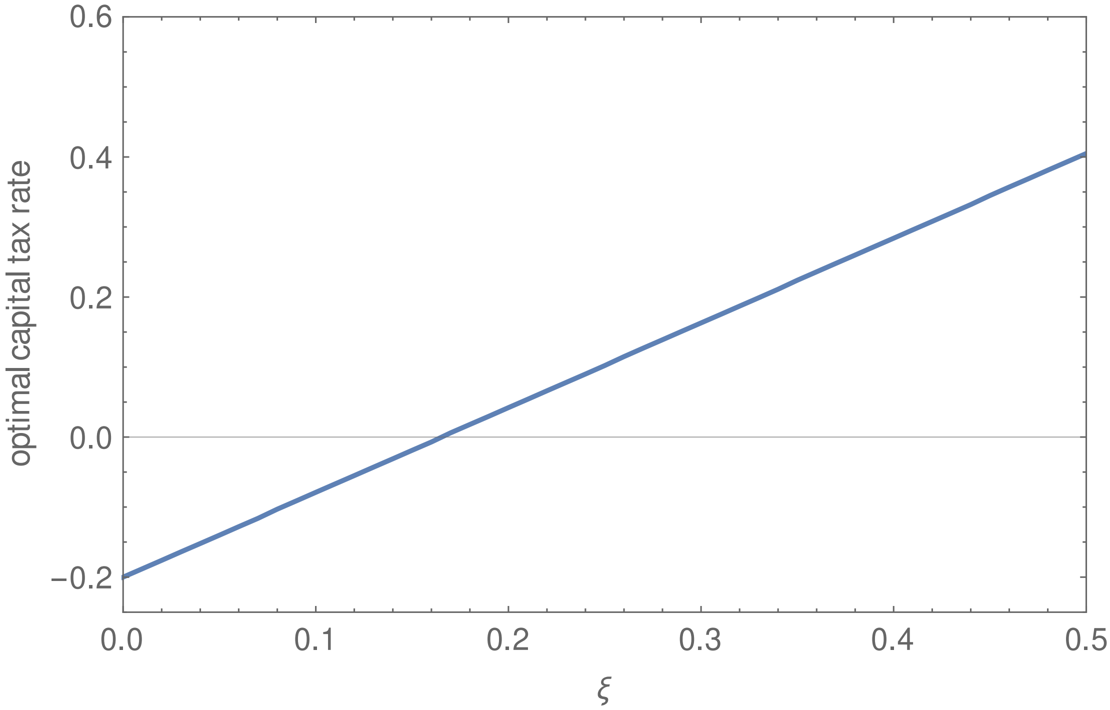

Figure 6: The optimal capital tax rate and the government size

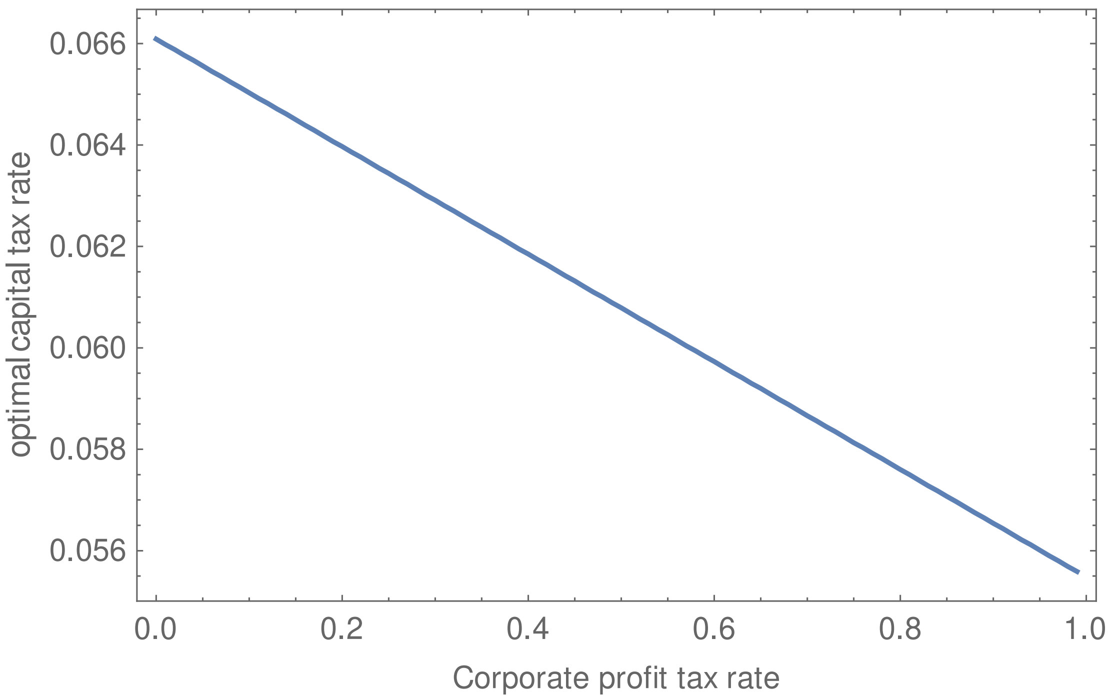

Figure 7: The optimal capital tax rate with exogenous corporate profit tax and endogenous labor income tax

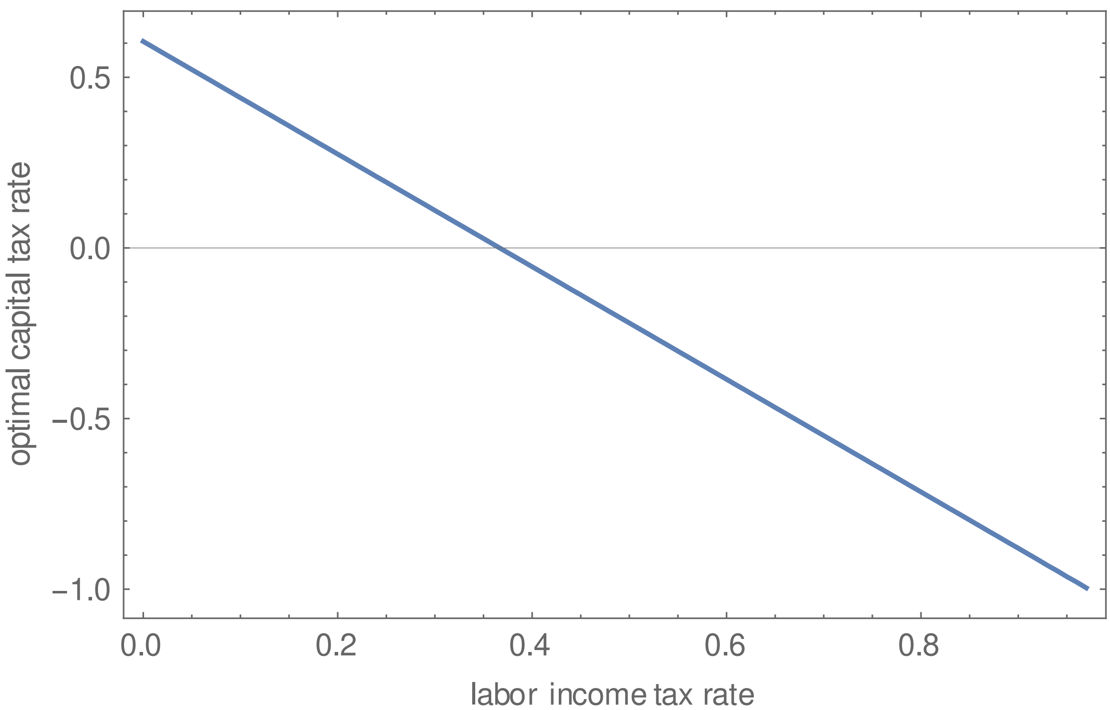

Figure 8: The optimal capital tax rate with exogenous labor income tax and endogenous corporate profit tax

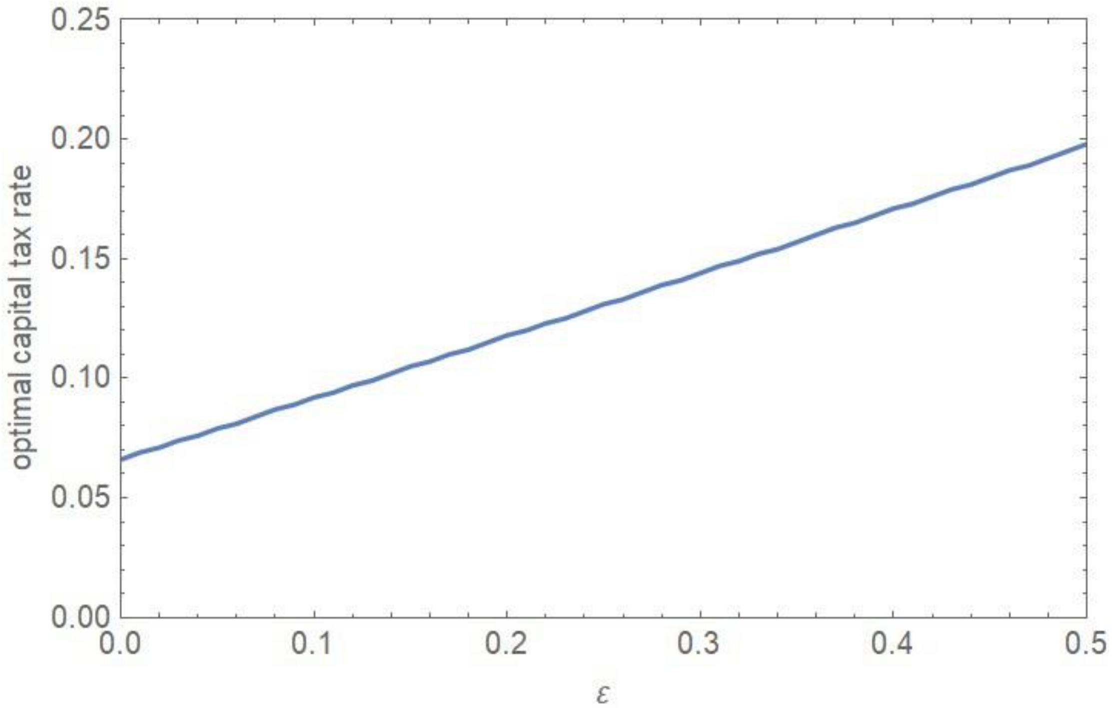

Figure 9: The optimal capital tax rate and the progressivity of labor income tax

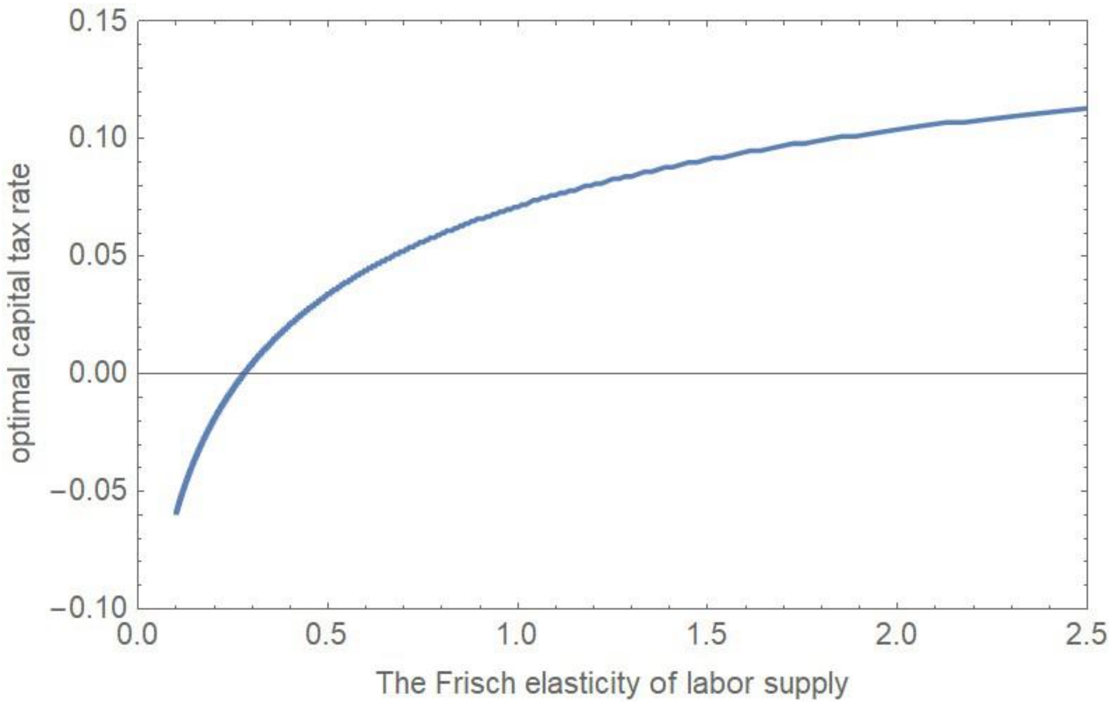

Figure 10: The optimal capital tax rate and the Frisch elasticity of labor supply

Table 1: Benchmark Parameterization

Definition	Parameter	Value	Source/Target
Labor income share	$1-\alpha$	0.6	Andolfatto et al. (2008)
Discount rate	β	0.05	Chu and Cozzi (2018)
Population growth rate	n	0.011	Conesa et al. (2009)
Initial capital tax rate	τ_K	0.3	Carey and Tchilingurian (2000)
Government size	ζ	0.22	Data
Leisure preference	χ	1.619	Total hours worked = 1/3
Stepping on toes effect	λ	0.2	Sequeira and Neves (2020)
Substitution parameter	ρ	2.2727	Monopolistic markup = 1.1
Standing on shoulders effect	ϕ	0.9837	Output growth rate = 2%
Size of innovation cluster	ψ	0.25	Comin (2004)
Capital depreciation rate	δ	0.0318	Capital-output ratio = 2.5

Table 2: Leisure preference and the optimal capital tax rate

	Targeted work-leisure allocation	Calibrated leisure preference	Optimal capital tax rate
High-work-time case	1: 1	$\chi = 0.797$	−0.4%
Benchmark	1: 2	$\chi = 1.619$	6.6%
Low-work-time case	1: 3	$\chi = 3.403$	11.0%