

# Optimal Tax Policy and Endogenous Growth through Innovation\*

Till Gross<sup>†</sup>  
Carleton University

Paul Klein<sup>‡</sup>  
Stockholm University

First draft: April 1, 2021  
This version: March 20, 2022

## Abstract

We investigate optimal tax policy in a Romer-style endogenous growth model. We derive formulas for the optimal tax rates on capital, labour, and innovation on a balanced growth path. We compute the balanced growth path and the transition to it with optimal policy for a range of parameter values. We find that capital should be taxed in the short run, but be paid its marginal product in the long run. The returns to innovation and production labour, on the other hand, should always be lower than their marginal products. Whether the resulting taxes on innovative activity should be positive or negative depends on (a) the extent of government spending needs, (b) the importance of innovation externalities and (c) the market power of patent holders. The welfare gains from optimal policy are much larger than in a comparable exogenous growth model.

*JEL Classification:* H21, E62, O3

*Keywords:* Dynamic Optimal Taxation, Ramsey Taxation, Innovation, Endogenous Growth, Technological Change

---

\*We thank Sergio de Ferra, Anastasios Karantounias, Florian Scheuer, Gustavo Ventura and seminar participants at Carleton University for their helpful comments and questions.

<sup>†</sup>Mailing address: Department of Economics, Carleton University, C-870 Loeb Building, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada. Email: [till.gross@carleton.ca](mailto:till.gross@carleton.ca).

<sup>‡</sup>Mailing Address: Department of Economics, Stockholm University, SE-10691 Stockholm, Sweden. Email: [paul.klein@ne.su.se](mailto:paul.klein@ne.su.se).

# 1 Introduction

In this paper, we explore the implications of optimal tax policy in the Ramsey (1927) sense in an environment similar to that of Romer (1990). Our first goal is to determine what optimal policy looks like in that environment. Our motivation for investigating this aspect of Romer’s model is that it is an environment where the ultimate engine of growth is endogenous innovation, innovation that responds to incentives. These incentives can to a large extent be influenced by policy, meaning that policy becomes an important determinant of long-run growth. This in turn implies that there is potentially more at stake than there is in an exogenous growth model à la Ramsey-Cass-Koopmans, and so the optimal policy choice may become more important. As Lucas (1988) famously said, “once one starts to think about [the determinants of long-term growth and its consequences for human welfare], it is hard to think about anything else.” Our second goal, then, is to determine the extent to which growth driven by endogenous innovation raises the stakes associated with tax policy.

To get a sense of how important growth is for welfare, consider a simple back-of-the-envelope calculation of the welfare benefits of a permanent increase in the growth rate. Suppose that consumption sequences are ranked according to the following utility function,

$$U = \sum_{t=0}^{\infty} \beta^t \ln c_t.$$

Now compare two consumption paths which are identical, except that one grows (faster) by a factor  $\Gamma$  in every period. By what factor  $\Delta$  must consumption in every period increase in order to make the other path as attractive as the faster growing one? Straightforward calculations reveal that

$$\Delta = \Gamma^{\frac{\beta}{1-\beta}}. \tag{1}$$

As an example, suppose every period is a year, and  $\beta = 0.98$ .<sup>1</sup> An increase in the

---

<sup>1</sup>This is the appropriate discount factor if the annual growth rate is 2% and the approximate after-tax rate of return is 4% per year, as documented by McGrattan and Prescott (2003).

growth rate by 0.1 percentage points, say from 2% to 2.1%, is then equivalent to a roughly 5% permanent level increase. This little calculation gives us a sense of what may be achieved through policy reform in the context of an environment, such as ours, where growth is endogenous in a way that responds to policy.<sup>2</sup> The point here is that even small changes in growth—including ones small enough to be hard to detect statistically—have large effects on welfare.

We base our model environment on that of Romer (1990) partly because it is one of the most well-known models of endogenous growth but also because we find the focus on innovation, which in turn is endogenously generated from skilled labour, particularly compelling. Moreover, as we detail in Section 2, optimal policy in such an environment has not been studied before. This claim may sound implausible in view of the existence of several papers that appear to do exactly that. However, upon closer examination, it turns out that those who have studied Ramsey optimal taxation in models based on Romer (1990), have in fact modified the model environment in such a way as to reduce them to versions of the Rebelo (1991) or “AK” model. The point is that, in these modified environments, the ultimate engine of growth is constant returns to cumulable factors. (It is then not very important whether we call these cumulable factors physical capital, human capital or ideas.) The basic properties of optimal policy in such an environment was studied already in Jones et al. (1993) and Jones et al. (1997). What they found was that, in the absence of externalities, all taxes would go to zero in the long run. What we do, by contrast, is to adhere more closely to the essential features of Romer (1990), and, as a result, our findings are quite different from those of Jones et al. (1993) and indeed from those of more recent papers in the area as detailed in Section 2. Nevertheless, our model is not quite identical to that of Romer (1990); we modify his model in the following four important ways for reasons that we find compelling.

---

<sup>2</sup>The same basic point is found in King and Rebelo (1990), who compared the welfare effects of tax changes in a Rebelo (1991) model (also known as the “AK” model) with those in an exogenous growth model. They concluded, as we do, that there is essentially an order of magnitude difference. That is not to say that we are merely expressing the conventional wisdom. Indeed, Lucas (1990) found that the growth and welfare effects of optimal tax reform in the context of the Lucas (1988) model, where long-term endogenous growth is driven by accumulation of human capital, are quantitatively trivial.

First, we assume that innovation consists of coming up with new types of intermediate goods, not new types of capital. The purpose of this modification is to enable us to distinguish between taxing capital and taxing innovation.

Second, we assume that innovative activity is associated with an externality that might be called a crowding effect. Specifically, how many new ideas an individual inventor comes up with increases linearly in that inventor's effort, but decreases in the aggregate amount of innovative effort. The rationale for this specification is the notion that certain ideas are, as it were, "in the air", just waiting to be invented, so that innovative effort is to some extent directed at being the first to come up with something as opposed to coming up with something that would not have been invented otherwise. A single parameter determines the extent to which innovative activity is productive as opposed to mere rent-seeking.

Third, patents (exclusive use to ideas) are assumed to expire after one period (which we think of as about 20 years) rather than to last forever, as they do in Romer (1990). We think this assumption is a reasonable one from an empirical standpoint. In this context we emphasize that, as in Jones (2019), we do not draw a sharp distinction between formal R&D on the one hand and entrepreneurial activity on the other. Consequently we think of patent duration not necessarily only in a narrow legal sense but as the time it takes for someone to copy an idea, whether patented or not.

In addition, there are two compelling model-driven reasons for our assumption of one-period-only exclusive rights to ideas. One is that it implies that the initial stock of ideas cannot be taxed (because nobody owns it), so (in contrast to capital) there is no temptation on the part of the government to confiscate it. Another is that it renders the optimal taxation problem a bit more tractable.

Fourth, like Romer (1990), we assume that labour is employed in innovation as well as in ordinary production, but, unlike him, we assume that each type of labour (innovative and ordinary) is supplied elastically. In our model, innovative labour can be used only to come up with new ideas, whereas ordinary labour can be used only for production. In Romer (1990), human capital is flexible in that it can be used

for either production or innovation, so as to make innovation activity endogenous; with our assumption of endogenous labour supply, this flexibility is not necessary. Furthermore, distinguishing between production labour and innovation labour allows us to make a conceptually clean distinction between taxing labour employed in production and taxing innovation.

Our main results are the following. First, we show analytically that, in the long run, capital should be paid its marginal product, while innovative activity and production labour should be paid less than their marginal products. As in Chamley (1986) and others, the initial capital stock should be taxed heavily. Since optimal taxation seeks to tax endowments (as opposed to effort/sacrifice) and there is no endowment of capital besides the initial one, long-run capital taxes are not used for the purpose of raising revenue (as opposed to correcting for distortions). At the same time, there is an endowment of time for production labour and an endowment of time for innovation in each period; therefore, the related economic activities will be taxed for the purpose of raising revenue in every period.<sup>3</sup>

Second, it is ambiguous whether the resulting taxes on profits from patents and, by implication, innovative activity, should be positive or negative. Due to the existence of externalities, innovative activity may be paid less than its marginal product and still face negative taxes. Whether innovation should be taxed or subsidized depends on (a) government spending needs (the desire for government spending), (b) the strength of the innovation externality, and (c) the degree of substitutability between different intermediate goods. The higher the desire for government spending, and the stronger the negative intra-period externality associated with innovation, the higher the optimal tax on innovation. These results are immediately intuitive.

Meanwhile, the elasticity of substitution between intermediate good varieties matters in a surprising way. The lower this elasticity of substitution, the higher the market power of firms that own patents and the higher their profits. Casual intuition would suggest that profit/patent taxes should be higher the higher the market

---

<sup>3</sup>These endowments have no fixed upper bound, but their use is associated with a convex cost. In this sense, labour supply involves both an endowment and effort/sacrifice.

power, but we find the opposite: a higher elasticity of substitution implies higher optimal profit taxes. To understand this, it may be helpful to note that the quantity supplied by a firm with market power is independent of the profit taxes; profits are still maximized no matter how much they are taxed. Moreover, the social gains from a new variety, the extent to which the final good is produced more efficiently, decreases in the elasticity of substitution between intermediate good. This becomes immediately clear in the limit, with perfect substitutes, where a new variety does not improve final good productivity at all. In a model of exogenous innovation, patents should be taxed at 100 percent regardless of the degree of substitutability since profits from patents are pure rents. In our dynamic model of endogenous innovation, however, the profits are not pure rents, but rather a reward for innovative activity. A lower elasticity of substitution implies larger social gains from innovation. Since these gains are not fully captured by inventors—patents expire after one period, but the productivity gains remain forever—patents should be taxed less the less substitutability there is.

Third, capital accumulation should be subsidized in the long run. This is a straightforward implication of increasing returns. If capital and labour were to receive their marginal products, this would more than exhaust output. So they are paid less than their marginal products, implying, in particular, that capital receives a market return that is strictly less than its marginal social return. As in Chamley (1986) and others, capital accumulation should not be distorted in the long run, which in our case implies a negative corrective tax. Meanwhile, there are many potential reasons, not present in our model, why optimal capital taxes may be positive; for instance, Conesa et al. (2009) provide a few of them. Our result is therefore not a sufficient argument for advising policy-makers that capital taxes should be negative rather than zero or positive. Rather, what our finding suggests is that if endogenous innovation is an important engine of growth, then that provides a reason to tax capital less than if endogenous innovation is unimportant for growth.

Finally, we emphatically confirm the view that long-term growth rates have an enormous impact on welfare, and that therefore optimal policy has the potential (if the

status quo is not optimal) to raise welfare by much more in a model of endogenous innovation than in an exogenous growth model. The welfare effects are greater, (i) the further the initial innovation taxes are from the ones on the optimal balanced-growth path (BGP), (ii) the lower the elasticity of substitution between different intermediate goods, and (iii) the lower the intra-period innovation externality. In our baseline calibration, optimal policy leads to welfare gains corresponding to a 6.0% permanent increase in public and private consumption,<sup>4</sup> whereas it would be 1.8% in the corresponding exogenous growth model. With different initial conditions and parameter values, the welfare gains can exceed 100%. Our model thus suggests that the stakes from optimal policy are indeed much higher in the presence of endogenous growth driven by innovation, and potentially an order of magnitude greater. Note that these welfare gains do *not* rely on an initial confiscation of capital because we assume that the initial capital tax rate is fixed and the time periods are so long that there is little temptation to impose a confiscatory rate in the second period either. Meanwhile, because exclusive rights to ideas only last for a single period and that new ideas can be used immediately (in the same period they were invented), there is no initial stock of ideas to tax.

Our paper is organized as follows. Section 2 situates our work in the context of the existing literature. Section 3 presents our model environment. Section 4 characterizes optimal policy, while Section 5 considers a calibrated version of the model and describes some quantitative features of the solution. Section 6 concludes.

## 2 Related Literature

Our contribution's place relative to the existing literature should be seen in light of the two main objectives of our paper that we highlighted in the Introduction: (i) we aim to understand the tradeoffs between innovation, labour in production, the accumulation of capital, and government spending needs, both in the short and in the long run, that determine optimal fiscal policy; and (ii) we want to evaluate

---

<sup>4</sup>We measure the welfare gain as the permanent percentage level change in both private and public consumption required to make an agent indifferent between that change and the reform. Had we given the compensation in terms of private consumption only, the number would be correspondingly higher. Specifically, about 8.5 percent rather than 6.0 percent.

the welfare effects of optimal fiscal policy reform, and specifically, whether these welfare effects are substantially larger than in an exogenous-growth framework.

In order to achieve the first goal, we need a model that clearly distinguishes the various choice margins, and a set of tax instruments such that for each margin there is exactly one instrument that directly affects that margin and nothing else; we call this a *complete* tax system.<sup>5</sup> In our model, the profit tax affects innovation, the labour tax affects production labour, and the capital tax affects the accumulation of capital. If one of these tax instruments were missing, say the profit tax, then optimal labour and capital taxes would have to take into account how they affect innovation. With profit taxes available, these effects do not have to be taken into consideration, since the incentives to innovate can be directly chosen through profit taxes.<sup>6</sup> By studying these three taxes jointly, our analysis is different from the rest of the literature of which we are aware.

We also emphasize that, in our model, government consumption is valued and endogenous rather than useless and exogenous. That turns out to matter much more in this context than it does in an exogenous growth model à la Cass (1965). One obvious reason for that is that if government consumption were completely exogenous (as opposed to, say, an exogenously fixed fraction of output), it would, generically, either vanish as a share of output in the long run, or eventually exhaust output. Less obviously but still crucially, if it were endogenous but useless, as in most of the existing related literature, then any policy that encourages growth would also automatically encourage useless government consumption, creating an artificial tradeoff between encouraging growth and avoiding waste.

---

<sup>5</sup>It is not obvious that a distinction between various forms of income can be drawn cleanly in practice. Indeed, the issue is the subject of much legal controversy. In 2017, 85 venture capitalists in Sweden lost a highly publicized legal case against the government and were required to classify a large proportion of their earnings as labour income rather than capital income. See, for example, <https://www.realtid.se/helena-stjernholm-forlorar-skattemaal-igen>. Correia (1996) and Reis (2011) consider incomplete tax systems.

<sup>6</sup>In particular, the incentive to innovate is given by the tradeoff between leisure and the net profits from having a patent. Labour and capital taxes affect gross profits, but profit taxes can always be chosen in such a way to obtain any net profits desired, independent of the gross profits (as long as these are strictly positive). Note, however, that a complete tax system does not imply that the solution is Pareto optimal or first best.



Furthermore, in a dynamic model, taxes have distinct incentive effects in the short and the long run, so allowing for time-varying tax rates is important. This is true even if one wants to consider only long-run (or only short-run) taxes. For instance, consider capital taxes in Chamley (1986). The government would like to heavily tax the initial capital endowment, but not tax capital in the long run at all. If the government could only tax capital at one time-invariant rate, then it would choose a moderate but strictly positive tax rate. This may or may not be policy relevant from a practical standpoint, but from the perspective of understanding the incentives involved in taxing capital, allowing only for time-invariant taxes would paint a misleading picture. By maintaining the assumption in Chamley (1986) that tax rates may vary over time, our approach stands out from the related literature.

When it comes to our second goal, it is important to point out that welfare analysis in a dynamic model requires consideration of both the short and the long run. Consequently, it is compulsory in this context to compute not only the BGP *but also the transition to it*. (The only exception is if the model is such that the economy is always on a BGP. This is true in the Rebelo (1991) model, but emphatically not in the Romer (1990) model.) For instance, consider yet again the model in Chamley (1986). There, a higher capital stock in the long run is desirable, but it comes at the cost of foregone consumption in the short and medium run. Therefore, a welfare analysis that only compares steady states would be incorrect. By computing fully-fledged transitions (as opposed to, say, considering only small perturbations from the BGP), we differ from all of the related literature on optimal taxation in innovation-driven endogenous growth models that we are aware of.

Another significant difference to the rest of the literature is the co-existence of monopolistic and competitive firms in the intermediate goods sector of our model—all other models have just one type of firm in the intermediate goods sector. While this feature is not essential for the questions we want to answer, we find it realistic, and it allows us to have finitely lived patents, limiting the severity of the time-inconsistency problem without having to make awkward assumptions about how

market power is passed on to another firm after a patent expires.<sup>7</sup>

Jones et al. (1993) analyze optimal taxation in the context of an endogenous growth model à la Lucas (1988), i.e. a model where growth is endogenous as a result of constant returns to cumulable factors. Because of their focus on the accumulation of human capital as opposed to innovative activity as the engine of growth, their policy prescriptions are quite different from ours. What they find is that, in the long run, both physical and human capital should not be taxed. The reason is that in their model, there are essentially only two endowments, the initial stocks of physical and human capital. These initial endowments should be taxed, leaving the long-run taxes (in the absence of externalities) at zero. By contrast, what we find is that while physical capital accumulation should not be distorted in the long run, both labour and innovative activity should be taxed in the sense that their after-tax returns are less than their marginal products. This is because there is an endowment of time for labour and innovation in each period, beyond the initial capital endowment.

Two more recent related papers are Croce et al. (2017) and Long and Pelloni (2017). Both share the following feature: Even though the model framework is inspired by Romer (1990), it differs crucially in that next-period innovation and current-period intermediate goods production both use the current-period final good as the only input. This implies that the model is essentially of the Rebelo (1991) type, with constant returns to cumulable factors, implying the absence of any transitional dynamics. Because returns to cumulable factors and labour jointly are increasing, factor prices are not equal to marginal products, leading to a distortion that may require correction through taxes or subsidies. In particular, the fact that the final good is an input into the production of intermediate goods, and vice versa, creates a multiplier

---

<sup>7</sup> The fact that our model has both monopolistic and competitive firms leads to an intratemporal misallocation of resources in the sense that there is less output of patented varieties relative to the varieties whose patents have already expired. We do *not* assume that there exists a tax instrument that can correct this inefficiency, for example a subsidy to monopoly output. An empirical motivation for this assumption is that it seems infeasible to implement such a subsidy in practice, especially since we view monopolistic firms not necessarily as literal patent holders, but more generally as entrepreneurs with a business idea that is sufficiently new that it hasn't been successfully copied yet. Moreover, the absence of such a tax instrument does not affect the optimal tax formulas that we find, with the sole exception of a modification to the elasticity of output with respect to innovation in Equation (29) on page 26 that would be required if we had that extra tax instrument.

effect that is external to individual firms and workers.

In both Croce et al. (2017) and Long and Pelloni (2017), government consumption is wasteful and equal to an exogenously fixed share of output, so there is a negative externality associated with growth, giving rise to a rationale for taxing capital (or, equivalently, profits or innovation, since neither model distinguishes between capital and innovation).<sup>8</sup> The output multiplier from intermediate goods implies that the marginal product of capital is higher than its private returns, giving rise to a countervailing reason to *subsidize* capital. In Croce et al. (2017), but not in Long and Pelloni (2017), there is an optimal subsidy per unit of intermediate goods, so this effect disappears.

In Croce et al. (2017), the specification of (negative) congestion effects and (positive) spillovers in innovation is such that they are hardwired to be of the same magnitude. Meanwhile, congestion is assumed to happen contemporaneously, whereas spillovers occur with a one-period lag. Consequently, as a result of discounting, the *net* externality associated with innovative activity is necessarily negative. In Long and Pelloni (2017), there are no innovation externalities, but the authors only allow for time-invariant taxes, so capital taxes are also used to partly expropriate the initial capital stock.<sup>9</sup> Annicchiarico et al. (2020) is in essence highly similar to Long and Pelloni (2017) and Croce et al. (2017), but it contains a negative “business stealing” externality associated with innovation.

One important feature of Croce et al. (2017) is aggregate shocks, which allows them to perform policy analysis over the business cycle. In this respect they stand out from the rest of the literature, including this paper, where the model is deterministic.

Cozzi (2018) aims to provide a credible quantitative assessment of the magnitudes

---

<sup>8</sup>Long and Pelloni (2017) have an extension with valued government consumption, but since they still have lump-sum transfers as a fixed fraction of output, this reason to tax capital remains.

<sup>9</sup>Their main result, that optimal capital taxes are positive, therefore does not stem from considering endogenous growth. On the contrary: If government spending were a constant fraction of output and if one allowed only for time-invariant taxes in Chamley (1986), then capital taxes would of course be positive, too. Endogenous growth in Long and Pelloni (2017), with effectively constant returns to capital and a rate of return that is lower than the marginal product, actually pushes capital taxes to be negative.

of taxes, and therefore incorporates realistic features such as overlapping generations, idiosyncratic productivity shocks, and non-linear labour taxes. Innovation is modelled as in Howitt and Aghion (1998), an extension of Aghion and Howitt (1992) that includes capital.

Aghion et al. (2013) study optimal taxation in a model based on Aghion and Howitt (1992). They rule out profit taxes (or R&D subsidies as they call them) as infeasible. Taxes are forced to be time-invariant and the authors do not compute a transition. Capital and labour taxes both affect the growth rate, through a so-called market size effect: higher taxes imply lower profits for inventors and hence reduce the growth rate. (This effect would have been present in our model too if we had ruled out profit taxes.)

### **3 The Model Environment**

Our model is a modified version of the one presented in Romer (1990). A final good is assembled by competitive firms from a variety of intermediate goods, which are imperfect substitutes. Blueprints for new varieties of intermediate goods can be invented, and these new varieties are the source of endogenous growth. One important difference in relation to the Romer model is that patents for newly invented varieties do not last infinitely long, but only for one period (which represents 20 years in our model). That is, new varieties are produced by monopolistic firms, and already existing varieties are produced by perfectly competitive firms. This assumption is more relevant empirically, as we argued above, but also makes the model's policy analysis more tractable and better suited for numerical analysis. The point here is that, under our assumption, the tax on innovation in period  $t$  directly affects innovative activity in period  $t$  only, as opposed to innovative activity in all previous periods  $s \leq t$ . If in this respect we had stuck to the specification of Romer (1990), the first-order conditions would have been plagued by an overwhelming abundance of derivatives.

### 3.1 Final Goods Producers

There is a large number of perfectly competitive firms that produce the final good, which can be used for private consumption, investment in capital goods, and government consumption. A representative firm produces output  $Y_t$  according to the following production function:

$$Y_t = \left( \int_0^\infty Q_{z,t}^\sigma dz \right)^{1/\sigma}, \quad (2)$$

where  $Q_{z,t}$  is the quantity used of intermediate input  $z$ , which is a continuous indicator;  $0 < \sigma < 1$  is a parameter that governs the degree of substitutability between these inputs, and  $1/(1 - \sigma)$  is the elasticity of substitution. The representative firm aims to maximize its profits (which are zero in equilibrium, due to perfect competition and constant returns to scale). Profits are given by

$$\Pi_{F,t} = \left( \int_0^\infty Q_{z,t}^\sigma dz \right)^{1/\sigma} - \int_0^\infty p_{z,t} Q_{z,t} dz, \quad (3)$$

where  $p_{z,t}$  is the price of intermediate input  $z$ , which each final-goods producer takes as given. As in the original Romer model, not all of these potential varieties are available at any time period, and we set the price of any unavailable input to infinity so that it is not used at all. This means that the integrals in Equations (2) and (3), in spite of appearances, in fact all have a finite upper limit defined by the stock of ideas already invented. The first-order condition for any available input yields its inverse demand function:

$$p_{z,t} = Q_{z,t}^{\sigma-1} Y_t^{1-\sigma}. \quad (4)$$

### 3.2 Intermediate Goods Producers

Each variety of intermediate good can potentially be produced by a large number of identical firms and they produce output  $Q_{z,t}$  according to a Cobb-Douglas production function:

$$Q_{z,t} = K_{z,t}^\alpha L_{z,t}^{1-\alpha},$$

where  $K_{z,t}$  is capital used by producer of variety  $z$  and  $L_{z,t}$  is labour. Each firm maximizes its profits per period:

$$\Pi_{z,t} = \max_{Q_{z,t}, K_{z,t}, L_{z,t}} p_{z,t}(Q_{z,t})Q_{z,t} - r_t K_{z,t} - w_t L_{z,t} \quad (5)$$

$$\text{s.t. } Q_{z,t} \leq K_{z,t}^\alpha L_{z,t}^{1-\alpha}, \quad (6)$$

where  $r_t$  and  $w_t$  are the rental rate of capital and the wage rate for labour, respectively. Each firm takes these and the inverse market demand function derived above as given. Cost-minimization implies that marginal costs  $MC_{z,t}$  for all firms are

$$MC_{z,t} = \left( \frac{w_t}{1-\alpha} \right)^{1-\alpha} \left( \frac{r_t}{\alpha} \right)^\alpha. \quad (7)$$

For the measure of varieties that has been newly developed,  $\hat{Z}_t$  (see below), there is only one firm for each such variety that holds the patent and hence monopoly rights for the current period. For all other varieties, there is a large number of perfectly competitive firms producing each intermediate good, and, due to constant returns to scale, profits are zero.

For monopolistic firms profit maximization yields

$$Q_{z,t}^M = \left( \frac{\sigma}{MC_{z,t}} \right)^{1/(1-\sigma)} Y_t$$

and

$$p_{z,t}^M = \frac{MC_{z,t}}{\sigma}.$$

For competitive firms the market output and price are respectively

$$Q_{z,t}^C = \left( \frac{1}{MC_{z,t}} \right)^{1/(1-\sigma)} Y_t$$

and

$$p_{z,t}^C = MC_{z,t}.$$

### 3.3 Blueprint Producers

In Romer, this is called the research sector, in which blueprints are produced through human capital (skilled labour)  $H$ , and we keep this notational convention. However, one could also think of entrepreneurs developing new business ideas, and we will occasionally use this language. Using skilled labour (or perhaps entrepreneurial effort)  $h_t$ , each inventor or entrepreneur can develop a measure of new ideas  $\hat{Z}_t$  according to the function

$$\hat{Z}_t = \eta Z_t h_t H_t^{-\Lambda}, \quad (8)$$

where  $Z_t$  is the current stock of ideas,  $h_t$  is individual innovative effort,  $H_t$  is aggregate innovative effort,  $\eta > 0$  is a parameter that governs the productivity of the innovation process, and  $0 \leq \Lambda < 1$  is a parameter that determines congestion in innovation. When  $\Lambda = 0$  as in Romer, then there is no concurrent innovation externality at all, and the productivity of one innovator does not affect the productivity of the others at all. As  $\Lambda$  approaches one, there is a fixed measure of ideas developed in each period, and innovative activity becomes purely rent-seeking. In any case, the stock of current ideas can be used by all and there is no depreciation of ideas, so we have

$$Z_{t+1} = Z_t + \hat{Z}_t. \quad (9)$$

As in Romer, we assume that a patent is granted for each new blueprint, but unlike Romer, this patent expires after one period. Researchers auction off their blueprints and the accompanying monopoly rights to the highest bidder (or entrepreneurs sell their new businesses to large investors/corporations), for a price  $P_t$ . We assume that new blueprints can be immediately put to use. This means that the measure of available ideas in period  $t$  is the sum of the measure of old ideas  $Z_t$  and the measure of new ideas  $\hat{Z}_t$  so that, for instance, Equation (2), defining the output of the final goods sector, becomes

$$Y_t = \left( \int_0^{Z_t} (Q_{z,t}^C)^\sigma dz + \int_{Z_t}^{Z_t + \hat{Z}_t} (Q_{z,t}^M)^\sigma dz \right)^{1/\sigma}.$$

### 3.4 Private Households

There is a measure one of identical consumers. We focus deliberately on the case of identical consumers, so as to highlight the efficiency effects of optimal fiscal policy with growth through innovation. Naturally, fiscal policy also affects the distribution of income, and it is an interesting and important question to study the distributional effects of fiscal policy with growth through innovation. That is a different research question for future work, though. The representative consumer takes prices and taxes as given and maximizes the following lifetime utility function:

$$u = \sum_{t=0}^{\infty} \beta^t [u(c_t, l_t, h_t) + v(G_t)]. \quad (10)$$

$u(c_t, l_t, h_t)$  is a period utility function that is strictly increasing and concave in consumption  $c_t$ , and strictly decreasing in labour  $l_t$  and entrepreneurial effort  $h_t$ . The subjective discount factor is denoted by  $\beta \in (0, 1)$ . Government consumption is denoted by  $G_t$  and  $v(\cdot)$  is a strictly increasing and concave function, with  $\lim_{G \rightarrow 0} v_G = \infty$ . All lower-case variables are individual level variables, and upper-case variables are aggregates. Because the population is constant, it may as well be equal to one, and hence there is no need to use the notation  $g_t$  for per-capita government consumption. In equilibrium, we will have  $C_t = c_t$  etc.

We assume that preferences are consistent with balanced growth, as described in King et al. (2002). A household's budget constraint for each period  $t$  is:

$$c_t = (1 - \tau_{l,t})w_t l_t + P_t Z_t h_t H_t^{-\eta} + (1 + R_t)a_t - a_{t+1} \quad (11)$$

where the tax on labour is denoted by  $\tau_{l,t}$ , the rate of return on assets is  $R_t$ , and  $a_t$  are asset holdings. Initial asset holdings  $a_0$  are exogenously given.

Utility maximization with respect to private consumption  $c_t$ , labour supply  $l_t$ , entrepreneurial effort  $h_t$ , and next-period assets  $a_{t+1}$  implies the familiar labour-leisure trade-off and "Euler" equation characterizing the optimal trade-off between consumption in the current period versus the next (subscripts denote partial deriva-



tives, e.g.  $u_{c,t} = \partial u(c_t, l_t, h_t) / \partial c_t$ :

$$-u_{l,t} = u_{c,t}(1 - \tau_{l,t})w_t \quad (12)$$

$$-u_{h,t} = u_{c,t}P_t Z_t H_t^{-\eta} \quad (13)$$

$$u_{c,t} = \beta u_{c,t+1}(1 + R_{t+1}). \quad (14)$$

Equations (11), (12), (13), and (14), together with initial asset holdings and no-Ponzi-scheme conditions (which we leave out to avoid tedious notation) characterize consumer behaviour.

### 3.5 Investors

Investors collect savings on behalf of consumers and invest them in capital, blueprints, and government bonds  $B_t$  in such a way as to maximize their profits (which are zero in equilibrium due to perfect competition and constant returns to scale).<sup>10</sup> Capital returns and profits are taxed at rates  $\tau_{k,t}$  and  $\tau_{z,t}$ , respectively;  $\hat{r}_t$  is the net return on government bonds. The representative investor's profit maximization problem is

$$\max_{\{K_t, B_t, A_t, \hat{Z}_t\}} (r_t - \delta)(1 - \tau_{k,t})K_t + \hat{Z}_t((1 - \tau_{z,t})\Pi_{z,t} - P_t) + \hat{r}_t B_t - A_t R_t \quad (15)$$

$$\text{s.t. } A_t = K_t + B_t, \quad (16)$$

where  $A_t = a_t$ , as noted earlier,  $\delta$  is the capital-depreciation rate, and  $K_t$  is the total amount of capital rented out.

The first-order conditions imply the usual no-arbitrage conditions and determine

---

<sup>10</sup>Investors are not a strictly necessary element of our model, but they simplify the exposition and analysis.

the price of a blueprint:

$$R_t = (r_t - \delta)(1 - \tau_{k,t}) \quad (17)$$

$$R_t = \hat{r}_t \quad (18)$$

$$P_t = \Pi_{z,t}(1 - \tau_{z,t}). \quad (19)$$

Equations (16), (17), (18), and (19) characterize the investors' behaviour. We refer to Equation (16) as the asset-market clearing equation.

### 3.6 The Government

The benevolent government maximizes the utility of its representative consumer. It finances endogenous government consumption  $\{G_t\}_{t=0}^{\infty}$  through taxes on capital income and profits, as well as through labour income taxes. Capital taxes are announced one period in advance, so the initial capital tax  $\tau_{k,0}$  is exogenously specified. Because of the length of the period, this almost entirely removes the temptation to tax the initial endowment of capital. Governments may also issue one-period debt (subject to a no-Ponzi-scheme condition, which we omit here). Initial government debt  $B_0$  is exogenously given. We impose an upper limit of 100% on capital and profit taxes (although neither is binding in our numerical analysis). The government's per-period budget constraint is

$$G_t + B_t(1 + \hat{r}_t) = \tau_{k,t}(r_t - \delta)K_t + \tau_{z,t}\Pi_{z,t}\hat{Z}_t + \tau_{l,t}w_tL_t + B_{t+1}. \quad (20)$$

### 3.7 Market Clearing

We have already implicitly assumed market clearing for intermediate goods, government bonds, and patents. Asset-market clearing is also inherent in the investors' problem. What remains is capital market clearing

$$K_t = \int_0^{Z_t} K_{z,t}^C dz + \int_{Z_t}^{Z_t + \hat{Z}_t} K_{z,t}^M dz$$

and labour-market clearing

$$L_t = l_t = \int_0^{Z_t} L_{z,t}^C dz + \int_{Z_t}^{Z_t + \hat{Z}_t} L_{z,t}^M dz,$$

where  $K_{z,t}^C$  and  $L_{z,t}^C$  are, respectively, the capital and labour demand of a competitive firm, while  $K_{z,t}^M$  and  $L_{z,t}^M$  are the capital and labour demand of a monopolistic firm. From the intermediate firms' cost minimization (details of the derivations are in Appendix B) we obtain the market-clearing prices

$$r_t = \alpha(K_t/L_t)^{\alpha-1} \left( Z_t + \hat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma} \quad (21)$$

$$w_t = (1 - \alpha)(K_t/L_t)^\alpha \left( Z_t + \hat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma}. \quad (22)$$

We can write total output as a function of aggregate capital, aggregate labour, and the stock of ideas as

$$Y_t = \frac{\left( Z_t + \hat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{1/\sigma}}{Z_t + \hat{Z}_t \sigma^{1/(1-\sigma)}} K_t^\alpha L_t^{1-\alpha}.$$

Note that “total factor productivity” in our model is  $\frac{\left( Z_t + \hat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{1/\sigma}}{Z_t + \hat{Z}_t \sigma^{1/(1-\sigma)}}$ , a more complicated expression than in other models of endogenous growth. The reason is that the markets for some intermediate goods are monopolistic, while others are competitive, so there are different amounts produced of different varieties.

We can express a monopolist's profits as

$$\Pi_{z,t} = (1 - \sigma) \sigma^{\sigma/(1-\sigma)} \frac{\left( Z_t + \hat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma}}{Z_t + \hat{Z}_t \sigma^{1/(1-\sigma)}} K_t^\alpha L_t^{1-\alpha}. \quad (23)$$

Final-goods market clearing is implied by Walras' Law:

$$Y_t = c_t + G_t + K_{t+1} - K_t(1 - \delta).$$

In Appendix A, we characterize a decentralized equilibrium as a function of (arbitrary) government policy.

## 4 Optimal Policy

The government announces its policy at time zero and is able to commit perfectly. Households, firms, and investors react optimally to this, as described above. The government can perfectly foresee how the private sector will react to its policies, and so it can use private-sector decisions as choice variables if it includes the private-sector optimality conditions as constraints. The Lagrangian is

$$\begin{aligned}
\mathcal{L} = & \sum_{t=0}^{\infty} \beta^t \{u(C_t, L_t, H_t) + v(G_t)\} \\
& + \psi_t \left[ \tau_{k,t}(r_t - \delta)K_t + \tau_{z,t}\Pi_{z,t}\widehat{Z}_t + \tau_{l,t}w_tL_t + B_{t+1} - G_t - B_t(1 + R_t) \right] \\
& + \theta_t \left[ (1 - \tau_{l,t})w_tL_t + P_t\widehat{Z}_t + (1 + R_t)A_t - A_{t+1} - C_t \right] \\
& + \mu_t [(1 - \tau_{l,t})w_tu_{c,t} + u_{l,t}] \\
& + \lambda_t \left[ u_{c,t}P_tZ_tH_t^{-\Lambda}\eta + u_{h,t} \right] \\
& + \zeta_t [\beta u_{c,t+1}(1 + R_{t+1}) - u_{c,t}] \\
& + \gamma_t [(r_t - \delta)(1 - \tau_{k,t}) - R_t] \\
& + \Upsilon_t [\Pi_{z,t}(1 - \tau_{z,t}) - P_t] \\
& + \omega_t [A_t - K_t - B_t] \\
& + \Omega_t \left[ Z_tH_t^{1-\Lambda}\eta - \widehat{Z}_t \right] \\
& + \kappa_t \left[ Z_t + \widehat{Z}_t - Z_{t+1} \right] \\
& + \phi_t [1 - \tau_{k,t}] \\
& + \varphi_t [1 - \tau_{z,t}],
\end{aligned} \tag{24}$$

with the following set of choice variables:

$$\tilde{X} = \{C_t, L_t, H_t, K_t, A_{t+1}, G_t, B_{t+1}, \tau_{k,t+1}, \tau_{l,t}, \tau_{z,t}, R_t, P_t, \widehat{Z}_t, Z_{t+1}\}_{t=0}^{\infty}. \tag{25}$$

We substituted out  $\hat{r}_t$  and replaced it with  $R_t$ , thereby eliminating Equation (18) as a constraint. We also have to keep in mind that  $r_t$ ,  $w_t$ , and  $\Pi_{z,t}$  are functions of  $K_t$ ,  $L_t$ ,  $\hat{Z}_t$ , and  $Z_t$  as described in Equations (21), (22), and (23):

$$\begin{aligned} r_t &= \alpha(K_t/L_t)^{\alpha-1} \left( Z_t + \hat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma} \\ w_t &= (1-\alpha)(K_t/L_t)^\alpha \left( Z_t + \hat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma} \\ \Pi_{z,t} &= (1-\sigma) \sigma^{\sigma/(1-\sigma)} \frac{\left( Z_t + \hat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma}}{Z_t + \hat{Z}_t \sigma^{1/(1-\sigma)}} K_t^\alpha L_t^{1-\alpha}. \end{aligned}$$

We now proceed to derive optimal capital, labour, and innovation taxes on a BGP. While our numerical analysis computes both the transition and the BGP, we focus for our analytical derivations on the BGP. For the remainder of this section, we assume that the economy converges to a non-degenerate BGP, along which private and public consumption as well as capital grow at a constant positive rate. In Section 5 we are able to confirm that, for a wide range of empirically relevant parameter values, the economy does indeed converge to a non-degenerate BGP. We conclude that, for practical purposes, the phenomenon described in Straub and Werning (2020) does not arise in the context of our model. All proofs are relegated to Appendix C.

## 4.1 Optimal Capital Taxes

We start with a characterization of optimal capital taxes:

**Proposition 1** *If the economy converges to a non-degenerate BGP, then the modified golden rule holds. The capital share, net of taxes, of total income is given by*

$$\frac{(1-\tau_{K,t})(r_t-\delta)K_t}{Y_t} = \frac{\partial Y_t}{\partial K_t} \cdot \frac{K_t}{Y_t} - \delta \frac{K_t}{Y_t}. \quad (26)$$

By  $\partial Y_t / \partial K_t$  we mean, of course,

$$\frac{\partial F(Z_t, \hat{Z}_t, K_t, L_t)}{\partial K_t}$$

where

$$F(Z, \hat{Z}, K, L) := \frac{\left(Z + \hat{Z}\sigma^{\sigma/(1-\sigma)}\right)^{1/\sigma}}{Z + \hat{Z}\sigma^{1/(1-\sigma)}} K^\alpha L^{1-\alpha}$$

and similarly for other partial derivatives of  $Y_t$ .

Proposition 1 says that the private return to capital should be equal to the social return to capital. In other words, the post-tax share of capital in total income should be equal to the contribution of capital to total output, by which we mean the elasticity of output with respect to capital. Importantly, this result is independent of the relative value of government funds, i.e. it does not depend on the Lagrange multipliers  $\psi_t$  (associated with the government budget constraint) versus  $\theta_t$  (associated with the household budget constraint). The return to capital should be undistorted, in the sense that Equation (26) also characterizes the first best.<sup>11</sup>

**Corollary 1** *Capital taxes on a non-degenerate BGP are negative.*

Corollary 1 makes clear that the return to capital being undistorted does not imply zero capital taxes. Capital taxes are not used to raise revenues, though; capital tax revenues are merely incidental. Rather, capital taxes are used in a corrective way, to align the private and public returns to capital, such that the capital provision is efficient. Due to constant returns to scale in capital and labour and non-zero profits, the rate of return on capital in the absence of taxes is smaller than the marginal product of capital. Therefore, negative capital taxes are used to align these two rates.

The negative capital tax result fits in with a broader result: in a wide set of models, capital is special in the sense that if all factors of production can be taxed independently, then in the long run the after-tax share of capital should, just as in our model, equal the elasticity of output with respect to capital. This finding is a further generalization of the results in Atkeson et al. (1999), though the precise statement is not

---

<sup>11</sup>In this context, our definition of “first best” is the welfare-maximizing competitive equilibrium allocation when lump-sum taxes, but not good-specific taxes and subsidies, are available. See also Footnote 7 on page 10.

that the capital tax should be zero in the long run. It is that, in the long run, the capital allocation should not be distorted and that therefore capital should not be taxed *in order to raise revenue*. Of course it does not follow that the optimal (long-run) tax rate on capital is zero. Indeed in our framework it is negative. Depending on the details of the model framework, the capital taxes that implement efficient intertemporal choices may be positive or negative or zero. An example of optimal (long-run) zero capital taxes is Chamley (1986), while an example of (potentially) positive capital taxes is Erosa and Gervais (2002). In any case, such non-zero optimal capital taxes are Pigouvian (corrective) in nature, and independent of the government's need to finance public spending (in contrast to other taxes, see below).

What can explain this special nature of capital? Optimal taxation seeks to tax endowments. In our model, there is an endowment of labour in each period, an endowment of entrepreneurial effort in each period, but only one endowment of capital, the initial capital stock. A government that does not have access to lump-sum taxes will raise revenues by taxing labour and entrepreneurial effort in each period, but will only seek to tax the *initial* capital stock—in the long-run, capital taxes are thus not used to raise revenues.

This logic relies on a complete tax system, where each factor of production can be taxed independently. Taxes on labour income do not directly tax the time endowment for labour, otherwise they would be non-distortionary, but they are better targeted to tax the labour endowment than capital or profit taxes. Similarly, taxes on profits do not directly tax the time endowment for entrepreneurial effort, but they are better suited than other taxes. The tax system can be incomplete for two reasons: (i) one of the factors of production cannot be taxed, as in Correia (1996), or (ii) the government cannot distinguish between income from at least two of these factors of production, as in Reis (2011). If, for example, the government could not tax labour in our model, then it could use profit and capital taxes to indirectly tax the time endowment for labour. Or if the government could not distinguish between labour and capital income in our model, then the joint tax would be a compromise between the desire to raise revenues from the labour endowment and the desire to

not distort the long-run capital allocation.

## 4.2 Optimal Labour Taxes

**Proposition 2** *If the economy converges to a non-degenerate BGP and if cross-derivatives of the period utility function are identically equal to zero, the labour share, net of taxes, of total income is given by*

$$\frac{(1 - \tau_{L,t})w_t L_t}{Y_t} = \frac{\frac{\partial Y_t}{\partial L_t} \cdot \frac{L_t}{Y_t}}{1/\tilde{\psi}_t + \left(1 - \tilde{\theta}_t/\tilde{\psi}_t\right) (1 + \epsilon_{L,t})}, \quad (27)$$

where  $\tilde{\psi}_t := \psi_t/(\beta^t u_{c,t})$  and  $\tilde{\theta}_t := \theta_t/(\beta^t u_{c,t})$ , which are constant on a BGP, and  $\epsilon_{L,t} := u_{LL,t} \cdot \frac{L_t}{u_{L,t}}$  is the elasticity of the marginal disutility from working. In Appendix D we provide a characterization of the optimal labour share, net of taxes, of total income with non-zero cross-derivatives. Naturally, the government takes into account how an increase in hours worked affects the marginal utility of consumption (and thereby the incentives to work, innovate, and save) and the marginal utility of innovating. We think that incorporating these additional terms does not add much to the intuition behind our results, and so we will stick to the simpler case of zero cross-derivatives (and hence an intertemporal elasticity of substitution of one).

**Lemma 1** *If the economy converges to a non-degenerate BGP and if cross-derivatives of the period utility function are identically equal to zero, as long as lump-sum taxes are unavailable, then  $\psi_t > \theta_t$ , and*

$$1/\tilde{\psi}_t + \left(1 - \tilde{\theta}_t/\tilde{\psi}_t\right) > 1. \quad (28)$$

The left-hand side of Equation (27) is the after-tax labour share of total income, and it is equal to the contribution of labour to total output (or the elasticity of output with respect to labour), divided by a term that is always greater than one as long as lump-sum taxes are unavailable. The elasticity of the marginal disutility from working,  $\epsilon_{L,t}$ , is always positive, since  $u_{L,t} < 0$  and  $u_{LL,t} < 0$ . Together with Lemma 1, it follows that the denominator of the right-hand side of Equation (27) is greater than



one. Therefore, the net return to labour will always be less than its contribution to output, and it is in this sense that labour supply is distorted, unlike capital.

Also unlike capital, the labour allocation depends on the marginal excess value of public funds,  $\psi_t - \theta_t$ . Equation (27) suggests that the higher it is, and hence the lower  $\tilde{\theta}_t/\tilde{\psi}_t$ , the lower is the after-tax share of labour. All objects in Equation (27) are endogenous, so we cannot conduct traditional comparative statics, but it suggests a relationship between the after-tax share of labour and the marginal excess value of public funds that is confirmed in all our numerical analysis. When lump-sum taxes are available, then  $\tilde{\psi}_t/\tilde{\theta}_t = 1$ , and  $\tilde{\psi}_t = 1$ , so that the after-tax share of labour is equal to the contribution of labour to total output.<sup>12</sup>

When lump-sum taxes are not available, then  $\tilde{\psi}_t/\tilde{\theta}_t > 1$ , and the after-tax share of labour is lower than the elasticity of output with respect to labour; how much lower also depends on the elasticity of the marginal disutility from working,  $\epsilon_{L,t}$ : the greater this elasticity, the lower is the after-tax share of labour (subject to the same caveat regarding comparative statics as above). To understand this, note that if the marginal disutility from working were constant, and thus  $\epsilon_{L,t} = 0$ , then a lower labour supply, resulting from a lower net wage due to higher labour taxes, would not relax the constraint with Lagrange multiplier  $\mu_t$ , which is for the constraint regarding the household's optimal labour supply. When  $\epsilon_{L,t} > 0$ , however, then the marginal disutility from working decreases as a person works less (or the marginal utility from leisure decreases as the person consumes more leisure). This implies that higher taxes, which result in a lower labour supply, would indirectly relax the labour-supply constraint; the government takes this into account and consequently sets higher labour taxes, the larger the elasticity of the marginal disutility from working.

---

<sup>12</sup>With lump-sum taxes, the government can freely move resources between the household and the treasury, and therefore  $\psi_t = \theta_t$  and  $v_{G,t} = u_{c,t}$ . Moreover,  $\psi_t = \beta^t v_{G,t}$  and in that case  $\theta_t = \beta^t u_{c,t}$ , so  $\tilde{\psi}_t/\tilde{\theta}_t = 1$ , and  $\tilde{\psi}_t = 1$ .

### 4.3 Optimal Profit Taxes

**Proposition 3** *If the economy converges to a non-degenerate BGP and if cross-derivatives of the period utility function are identically equal to zero, the profit share, net of taxes, of total income is given by*

$$\frac{(1 - \tau_{Z,t})\Pi_t \widehat{Z}_t}{Y_t} = (1 - \Lambda) \cdot \frac{\frac{\partial Y_t / Y_t}{\partial \widehat{Z}_t / \widehat{Z}_t} + \frac{\beta \widehat{Z}_t / Z_t}{(1 - \beta)(1 + \widehat{Z}_t / Z_t)} \left( \frac{\partial Y_t}{\partial Z_t} \cdot \frac{Z_t}{Y_t} + \frac{\partial Y_t}{\partial \widehat{Z}_t} \cdot \frac{\widehat{Z}_t}{Y_t} \right)}{1 / \widetilde{\psi}_t + \left( 1 - \widetilde{\theta}_t / \widetilde{\psi}_t \right) (1 + \epsilon_{H,t})}, \quad (29)$$

where  $\epsilon_{H,t} := u_{HH,t} \cdot \frac{H_t}{u_{H,t}}$  is the elasticity of the marginal disutility from providing entrepreneurial effort. In Appendix D we provide a characterization of the optimal profit share, net of taxes, of total income with non-zero cross-derivatives.

In Equation (29) we can see that the after-tax share of profits in output is equal to the elasticity of current and future output with respect to an additional variety of intermediate goods (the numerator of the right-hand side), divided by a term similar to the one for the after-tax labour share, and multiplied by  $1 - \Lambda$ . We focus first on the term  $1 - \Lambda$ : When  $\Lambda$  approaches 1, and technical progress is independent of research activity, then all research effort is rent-seeking in nature, and the after-tax share of profits should be zero. The smaller this research externality, the larger the after-tax share of profits.

The numerator of the right-hand side of Equation (29) contains three terms: The first,  $\frac{\partial Y_t / Y_t}{\partial \widehat{Z}_t / \widehat{Z}_t}$  is the direct impact of a new variety on contemporary output produced by a monopolistic firm (expressed as an elasticity). The second term,  $\frac{\partial Y_t}{\partial Z_t} \cdot \frac{Z_t}{Y_t}$  is multiplied by  $\frac{\beta \widehat{Z}_t / Z_t}{(1 - \beta)(1 + \widehat{Z}_t / Z_t)}$ . An additional variety today results in one more variety produced by competitive firms in the next period, and in fact in all future periods. The third term,  $\frac{\partial Y_t}{\partial \widehat{Z}_t} \cdot \frac{\widehat{Z}_t}{Y_t}$  is also multiplied by  $\frac{\beta \widehat{Z}_t / Z_t}{(1 - \beta)(1 + \widehat{Z}_t / Z_t)}$ , and reflects the fact that an additional variety today makes producing additional varieties in the next period, and again in all future periods, easier; one should note that the production of new knowledge scales linearly in current knowledge. All together, the numerator captures the effects of researching an additional variety on current and properly

discounted future output.

Similar to the labour-tax share, Lemma 1 together with the fact that  $\epsilon_{H,t} \geq 0$  implies that the denominator of the right-hand side is always greater than one, and that hence the after-tax share of profits is always lower than the contribution of entrepreneurial effort to current and future output. When lump-sum taxes are available, then the denominator is equal to one, and entrepreneurial activity is not being “taxed;” when public funds are relatively scarce, and  $\psi_t > \theta_t$ , then the denominator is strictly greater than one, and entrepreneurial activity is being taxed, in the sense that the private returns to effort are lower than the social marginal product. However, there are three externalities: (i) the contemporary one captured by  $\Lambda$ , (ii) the intertemporal one that a patent lasts one period, whereas the new variety lasts forever, and (iii) the intertemporal one that future research builds on current research. This means that actual taxes on profits may be positive or negative, depending on whether the private return to innovation is greater than the social return or lower.

Using the same reasoning (and caveats) as for the labour-tax share, the after-tax share of profits is decreasing in the marginal excess value of public funds,  $\psi_t - \theta_t$ , and the elasticity of the marginal disutility from innovation  $\epsilon_{H,t}$ . The taxation of innovation thus follows the same logic as labour taxation and not capital taxation. What matters is not whether a factor of production has an intertemporal dimension to it: taxes on innovation clearly distort the intertemporal tradeoff. What matters is taxing endowments: there is one initial capital endowment, which calls for high initial taxes on capital, but then taxes in the long run should not distort the capital allocation; however, there is a labour time endowment and an innovation time endowment in every period, so these factors should be taxed in the short and the long run. As we show for capital and innovation, whether the government aims to tax a factor or not, the tax rates that implement the allocation may be positive, zero, or negative.

One factor that affects these actual tax rates, but that is not immediately visible in Equation (29), is the degree of substitutability between different types of intermediate goods, determined by the parameter  $\sigma$ . The elasticity of substitution is  $1/(1 - \sigma)$

so that intermediate goods are better substitutes the closer  $\sigma$  is to one. The terms  $\frac{\partial Y}{\partial Z_t}$  and  $\frac{\partial Y}{\partial Z_t}$  are a function of  $\sigma$ . Generally speaking, the higher  $\sigma$ , the lower these derivatives: a low substitutability implies that a new variety cannot be easily replaced by an already existing one, and hence a new variety increases final good productivity by more than with a high substitutability. If a higher  $\sigma$  implies that the right-hand side is lower, then the left-hand side needs to be lower, too. This can be achieved by higher profit taxes. As before, the qualifier holds that all variables change with  $\sigma$ ; however, our numerical analysis confirms this relationship between profit taxes and  $\sigma$ . The intuition is that the less substitutable intermediate goods are, the greater is the social gain from another invention. Because patents expire after one period, these gains are not fully captured by inventors. Hence patents should be encouraged more, i.e. taxed less or subsidized more, the less substitutability there is.

Based on the findings in Jones and Williams (1998), it is at least arguable that the most plausible case empirically is that innovation should be subsidized.

## 5 Numerical Results

In this section we evaluate the quantitative importance of optimal taxation on growth and welfare and conduct some numerical experiments.

### 5.1 Parameterization

We follow convention and set the parameter  $\alpha$  in the production function of intermediate goods to  $1/3$ . An important parameter is  $\sigma$ , which determines the degree of substitutability between different intermediate goods. As  $\sigma$  approaches one, these become perfect substitutes. We set  $\sigma = 6/7$ , which corresponds to an elasticity of substitution of 7, which fits within the range of estimates from the international trade literature; see for instance Anderson and van Wincoop (2004)[p.33], who write: "Overall the literature leads us to conclude that [the elasticity of substitution] is likely to be in the range of 5 to 10." One period is 20 years, which conforms roughly to the length of a patent life. This choice has several advantages: (i) it keeps the number of periods before convergence to a BGP low, and thus facilitates computation of the equilibrium; (ii) patents last exactly one period, which keeps both the

analytical and computational work more tractable, since only current profits affect the choice to develop new patents; (iii) most importantly, there is no incentive to tax the existing stock of ideas. We set the depreciation rate  $\delta$  to 0.8, which conforms to a yearly depreciation rate of about 8%. We assume that utility takes the form

$$u(c_t, l_t, h_t) + v(g_t) = (1 - g) \log(c_t) - e_L L^\epsilon - e_H H^\epsilon + g \log(g_t),$$

so that the intertemporal rate of substitution is equal to one. We set the Frisch elasticity of labour supply to 0.5, which corresponds to  $\epsilon = 3$ . A key parameter is  $\Lambda$ , which we set arbitrarily, but not unreasonably to 0.5. It seems very difficult to pin down what this parameter should be. Instead of trying to calibrate it to a specific, but in our opinion still arbitrary target, we conduct extensive sensitivity analysis regarding this parameter.

The remaining parameters,  $e_L$ ,  $e_H$ ,  $g$ ,  $\beta$ ,  $\eta$ , and the initial debt  $B_0$  and assets  $A_0$  are calibrated by computing a “pre-initial BGP”, which satisfies all of the government’s first-order conditions for optimality, except for capital and profit taxes (which we fix to 40% each) and next-period government debt. On a BGP, the latter condition is redundant, and conceptually speaking government debt is not pinned down on a BGP, as it is determined by initial conditions. We thus replace the FOC for debt by a calibration condition, that the debt-to-GDP ratio on the BGP is equal to its target. We set the following targets: (i) time worked in the production sector is  $L = 1/3$ , which is a normalization; (ii) time worked in innovation is  $H = 1/3 \cdot 1/9$ , since one tenth of the workforce in the United States is self-employed, which we take as a rough measure of innovators, but this is again a normalization; (iii) the ratio of government revenues to GDP on the pre-initial BGP is 30%; (iv) the yearly interest rate is 4% per annum, roughly corresponding to a value of 1.2; (v) the growth rate of the economy on the pre-initial BGP is 2% per year, which is equivalent to about 0.5; (vi) the debt-to GDP ratio on the pre-initial BGP is 0.03, which corresponds roughly to 60% on an annual basis, which is the ceiling of the European Stability and Growth Pact.

We then have 20 constraints and 20 variables from the modified government’s op-

timal tax problem (the constraints on maximum taxation are not binding on an interior BGP, so we do not include the constraints and the Lagrange multipliers); in addition, we have eight “target” equations, for capital and profit taxes, skilled and unskilled labour, the ratio of government revenues and government debt to GDP, the interest rate, and the growth rate of the economy. All together, this results in 28 equations and 28 variables, and we solve this using the Broyden method. Due to constant returns to scale in capital and labour, the initial amount of assets  $A_0$  follows from this, and does not require an extra target.

Our numerical procedure for the optimal policy is similar: with the parameters and initial conditions given, we compute the transition and BGP simultaneously. The optimal-taxation problem generally depends on initial conditions, and it is thus not possible to compute a steady state or BGP independently of the transition. We use the Broyden method to solve a system of 24 equations and 24 variables per period (the cap on profit taxes is never binding, so we do not include the constraint and Lagrange multiplier in this), for  $T = 25$  periods. This corresponds to a time frame of 500 years and is enough for (approximate) convergence to a BGP. We force variables at period  $T + 1$  to be equal to their counterparts in period  $T$ , adjusted for growth. We provide the details in our computational appendix.

## 5.2 Baseline Results

For our baseline calibration, we show our results in Figure 1. Capital taxes in period zero are at their exogenously imposed level, 40%, drop to 8.5% in the first period, and by the second period the convergence to the BGP level of -17.6% is almost complete. The reason why capital is not more strongly taxed in the first period is because the period length is twenty years, which means that the anticipation effects are very strong. Unlike in Atkeson et al. (1999) capital taxes do not immediately converge to their BGP level after one intermediate period, since optimal capital taxes also depend on growth: The difference between the private and the social returns to capital are a function of innovation activity. Notice that the economy does indeed converge to a non-degenerate BGP, meaning that the phenomenon described in Straub and Werning (2020) does not arise.

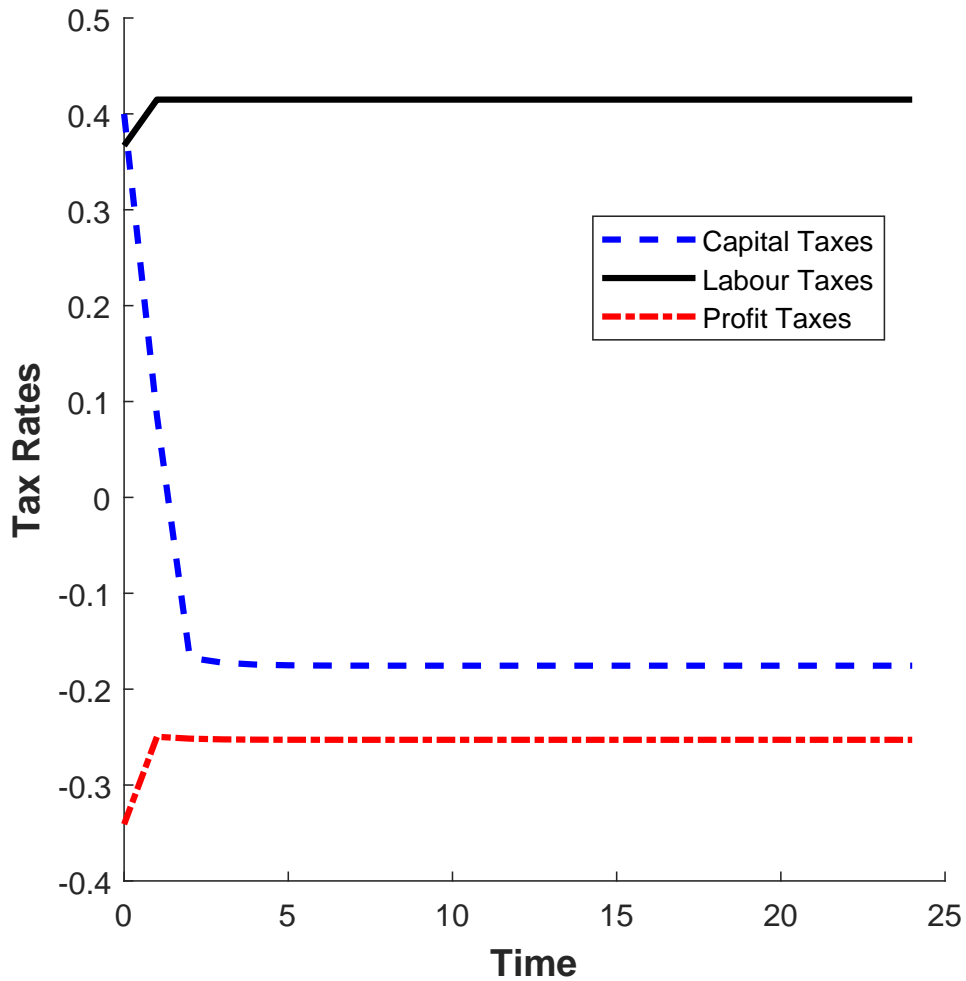


Figure 1: Optimal taxes over time in the baseline scenario.

Profit taxes are negative, reflecting a relatively low level of the intratemporal research externality and desire for public funds. Labour and profit taxes are at a lower level in period zero, and (almost) at their long-run level afterwards. Taxes other than those on capital are low initially to help tax the initial capital stock. Low taxes increase production and low government consumption, see Figure 2, encourages private consumption; this lowers in turn the marginal utility of consumption, and thereby decreases the value of the initial capital stock. We explore this in more detail in Gross et al. (2021). The government uses the funds from the taxation of the initial

capital stock to lower its debt and in fact accumulate assets.<sup>13</sup>

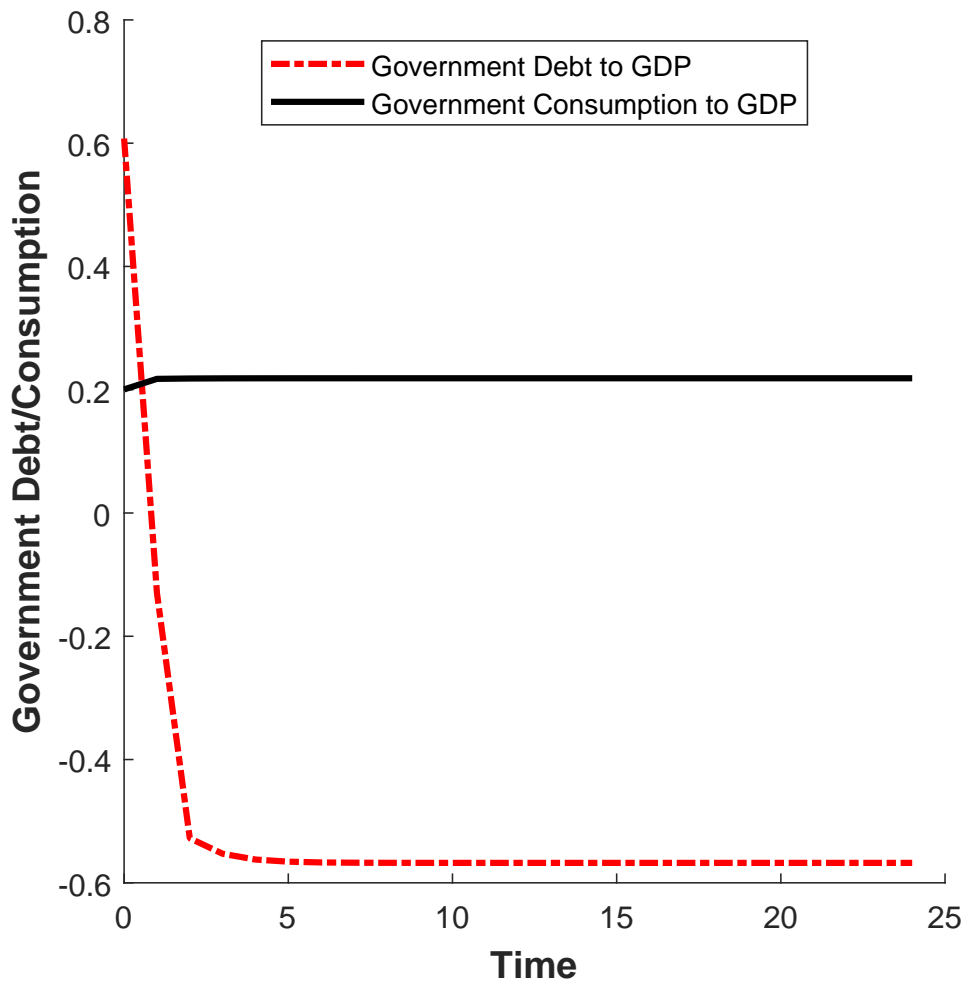


Figure 2: Optimal government consumption and debt, as a fraction of GDP, over time, annualized, in the baseline scenario.

### 5.3 Varying the Research Externality, Elasticity of Substitution, and Value of Public Funds

We now turn to examine the importance of our calibration choices for our results. As our theoretical work has shown, the intraperiod research externality  $\Lambda$  plays an

---

<sup>13</sup>This may in turn be growth-enhancing if government debt is a risk factor that affects innovation, see Croce et al. (2019) and Croce et al. (2021). On the other hand, there may be other reasons why governments would like to hold debt, see for instance Aiyagari and McGrattan (1998).



important role, as well as the elasticity of substitution between varieties  $1 - 1/\sigma$ , and the value of public funds  $\psi - \theta$ , which is mainly influenced by the parameter  $g$  and hence our calibration target for the ratio of government revenues to GDP. In the following exercises, we recalibrate all parameters. We start with the time path of optimal fiscal policy for a higher value of  $\Lambda$  and then turn to the innovation taxes on a BGP for various different combinations of  $\Lambda$ ,  $\sigma$ , and the calibration target for the ratio of government revenues to GDP.

As shown in Figure 3, the temporal pattern stays the same when we increase the value of  $\Lambda$  to 0.75 as compared to 0.5 in the baseline, but the profit taxes are now positive. Labour taxes are significantly lower, since the subsidies to research no longer have to be financed, and instead some revenues are generated from profit taxes. Capital taxes, on the other hand, are very similar to the baseline, with two minor differences: (i) the capital tax in period one is a bit lower at 5.4%, because the need to tax the initial capital stock is lessened, due to the revenues from profit taxes; (ii) capital taxes on the BGP are a bit higher, at -17.0%, since growth is lower, hence the fraction of monopolistic firms is lower, and thus the difference between the social and private returns to capital is smaller.

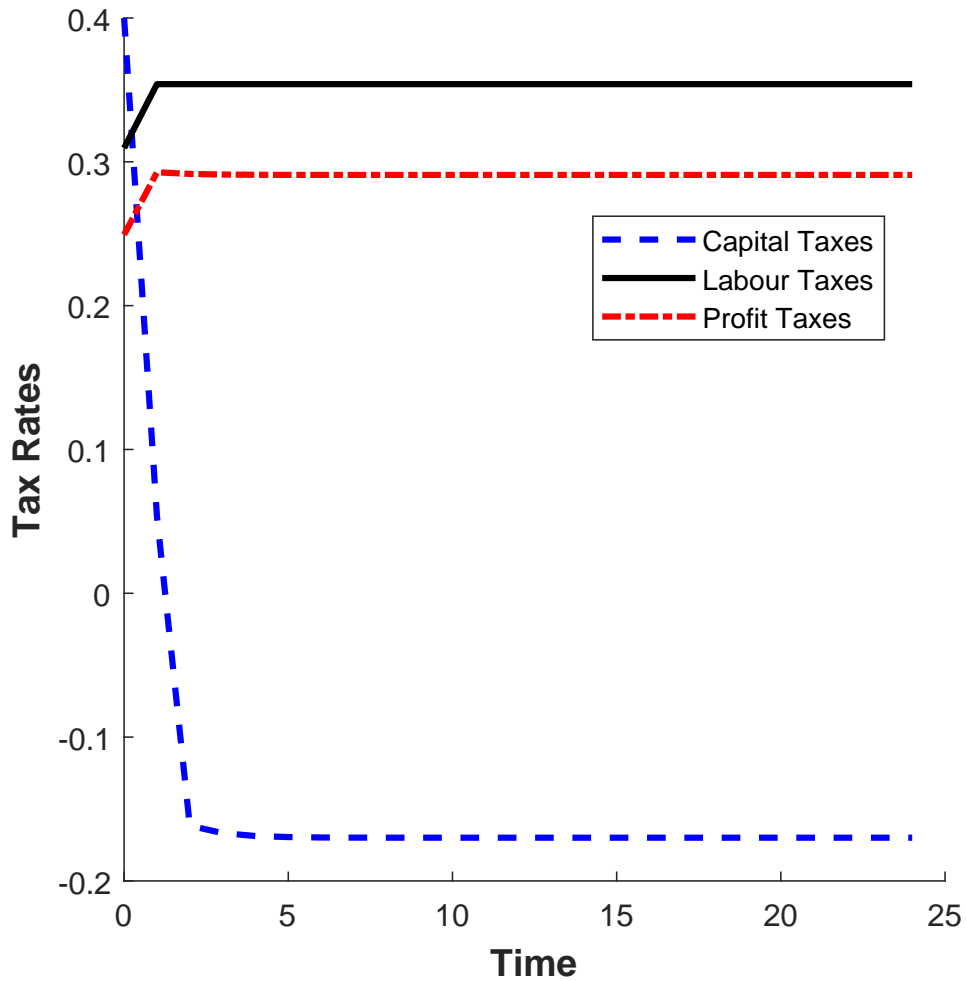
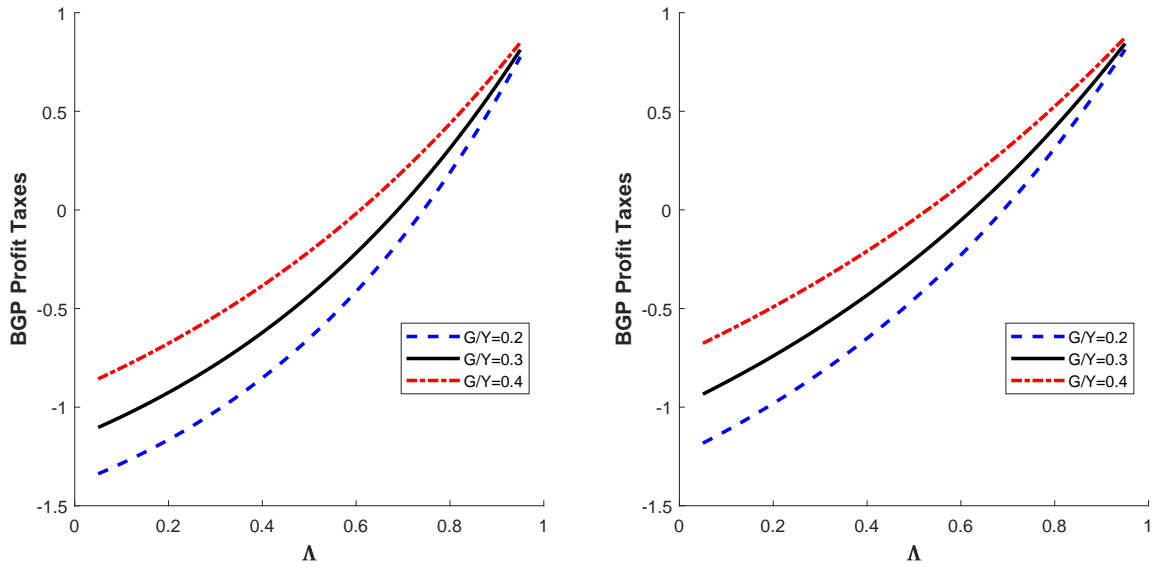


Figure 3: Optimal Taxes over Time with a High Lambda

In the following we provide two graphs that illustrate the importance of the value of public funds, the research externality  $\Lambda$ , and the degree of substitutability between varieties  $\sigma$ , which relates inversely to the degree of market power. Panel (a) of Figure 4 shows profit taxes on a BGP for three different levels of public spending as a function of  $\Lambda$  for a low  $\sigma$ , corresponding to an elasticity of substitution between varieties of 5, and Panel (b) shows it for an elasticity of 7.



(a) Low elasticity of substitution  $\sigma = 4/5$  (b) High elasticity of substitution  $\sigma = 6/7$

Figure 4: Profit Taxes on a BGP

The first observation we can make is that profit taxes are higher, the stronger the preference for public spending. This may seem like an obvious result, but it is worth repeating in this context that long-run capital taxes do *not* depend on this preference. The optimal capital taxes do vary, but this is because the difference between the marginal product of capital and the private returns to investors depends on the profits by patent-holders. These profits depend in turn on the degree of substitutability between varieties, but also the tax policy, in particular profit taxes. Capital taxes thus follow qualitatively the same pattern as profit taxes, but as our analytical derivations made clear, this same pattern is due only to the fact that it affects the difference between public and private returns to capital. Moreover, the quantitative impact is small compared to profit taxes: BGP capital taxes vary for all these different parameter values between -16.8% at their highest and -20.8% at their lowest, while profit taxes are -133.8% at their lowest and 87.4% at their highest.

What explains this difference? Both technology and capital accumulate over time, but profit taxes are taxing the time endowment of inventors/entrepreneurs, while

capital taxes tax the initial capital endowment; after the initial taxation of the capital endowment, capital taxes are only used to achieve the modified golden rule, and are thus independent of public spending preferences, subject to the same caveat as in the previous paragraph. Since there is no capital endowment beyond the initial one, there is no tradeoff between taxing an endowment and efficiency for capital taxes, and long-run capital taxes are purely used for maximum efficiency. For profit taxes, there is no initial endowment of patents to tax, but the time endowment of inventors can be taxed in every period, including on the BGP, at the expense of a reduced efficiency in research activity. Profit taxes thus depend on the preference for public spending, as there is always a tradeoff between taxing an endowment and efficiency – profit taxes are in this regard more akin to labour taxes than capital taxes.

The second observation is that the higher the negative intra-period research externality  $\Lambda$ , the higher are profit taxes. This is immediately intuitive. At the extreme, as  $\Lambda$  approaches one, the aggregate number of new varieties is independent of the time spent by inventors, and inventors are merely fighting over who patents it first. In this case, profits by inventors are economic profits and should be taxed away completely; if we had an exogenous growth model where some firms have market power (corresponding to the limit of  $\Lambda \rightarrow 1$ ), a tax on profits is a lump-sum tax. As the graphs illustrate, the desire for public spending is also less important, the higher  $\Lambda$  is; this is because profit taxes distort less the number of new varieties invented, the higher  $\Lambda$  is. At the limit of  $\Lambda \rightarrow 1$ , there is no distortion at all, and profit taxes of 100% are optimal, no matter what the public spending preferences are.

The third observation is perhaps a little counter-intuitive at first: profit taxes are increasing in the degree of substitutability between goods, which implies that as firms with patents have more market power and charge larger markups, profit taxes are lower. At first sight, this seems exactly the opposite of what a government should do: casual intuition would suggest that the higher the profits due to market power, the higher the profit taxes. In our model, a lower elasticity of substitution between varieties (and hence more monopoly power for patent holders) implies that a new

variety adds more to the final good than with a higher elasticity. At the limit, with perfect substitutes, a new variety does not improve final-good productivity at all. So far, we can thus state that a lower elasticity leads to more private gains for the inventor (through higher profits), but also to higher social gains (via increased productivity). Profit taxes are increasing in the degree of substitutability, because inventors reap the private profits for only one period, the twenty years of the patent length, whereas the benefits last forever.

## 5.4 Welfare Gains from Optimal Policy

We now explore the welfare effects of optimal policy for various parameter values and different pre-initial capital and innovation taxes. The pre-initial BGP is calculated in exactly the same way as described earlier for the calibration, which means that we recalibrate all the parameters in each of the scenarios that we are considering. As of some period  $t = t_0$ , the government may choose policy optimally (though the initial period's capital tax is given by history to avoid a confiscatory rate). Our welfare gains are computed by comparing the utility of the representative agent with an optimal fiscal policy to the utility if the economy stayed on the pre-initial BGP. We express these welfare gains as the permanent percentage level increase in private and public consumption required to make the pre-initial BGP achieve the same utility as the path implied by optimal policy.

In order to be able to make a reasonable comparison to the welfare gains from optimal policy one would obtain in an exogenous growth model, we consider an economy in which all firms are competitive, there is no skilled labour  $H$ , and in which the technological growth rate is fixed. For details, see Appendix E. The calibration for this economy follows the same logic and the same targets as for the endogenous growth model. In all of the following, the capital taxes are set to 40% on the pre-initial BGP.

We start with our baseline profit taxes of 40% on the pre-initial BGP and show the results in Figure 5. The welfare gains in an exogenous growth model are naturally constant as we vary the parameters  $\Lambda$  and  $\sigma$ , at 1.8%. In our endogenous growth

model, the welfare gains from optimal policy depend very much on the parameters. For lower levels of  $\Lambda$  and  $\sigma$ , the welfare gains are decreasing in  $\Lambda$  and  $\sigma$ , but then the relationship reverses. What can account for this non-monotonic relationship? It comes from the profit taxes on the pre-initial BGP: the closer these taxes are to the optimal profit taxes on the optimal BGP, the lower are the welfare gains from optimal policy. Profit taxes of 40% on the BGP are optimal when  $\Lambda$  is about 0.8, so that is roughly where we see the turning point.<sup>14</sup> That is also roughly the point where the role of  $\sigma$  reverses.<sup>15</sup> When profit taxes on the pre-initial BGP are suboptimally high, then the gains from lowering these to the optimal level and increasing growth is larger the lower  $\sigma$  is (since a lower  $\sigma$  implies larger gains from innovation). However, when profit taxes on the pre-initial BGP are suboptimally high, then the optimal tax reduces the growth rate compared to the pre-initial BGP and hence the welfare gains are smaller when  $\sigma$  is lower.

Another observation we can make is that in our endogenous growth model, the welfare gains from optimal policy are larger than in the exogenous growth model, and in some cases an order of magnitude larger. In our benchmark specification, with  $\Lambda = 0.5$  and  $\sigma = 6/7$ , the welfare gains are 6.0%. In the most optimistic case, when  $\Lambda = 0.05$  and  $\sigma = 4/5$ , the welfare gains are 26.3%, more than 14 times larger than with exogenous growth; at the same time, the most pessimistic case yields welfare gains that are in the vicinity of those in an exogenous growth model, 2.8%. This last result is to a large extent due to our assumption that there is no initial stock of patents to tax; in a model as in Romer (1990), where patents are infinitely-lived, there would be substantial welfare gains from expropriating the owners of existing patents.

---

<sup>14</sup>It is tempting to think that the turning point should be precisely where optimal profit taxes on the BGP are 40%, but that is not quite correct. Optimal profit taxes also depend on capital taxes and the level of debt, which differ on the optimal BGP as compared to the pre-initial BGP.

<sup>15</sup>Again, this is not precisely the point of reversal, as the optimal profit tax on the BGP also depends on the value of  $\sigma$ .

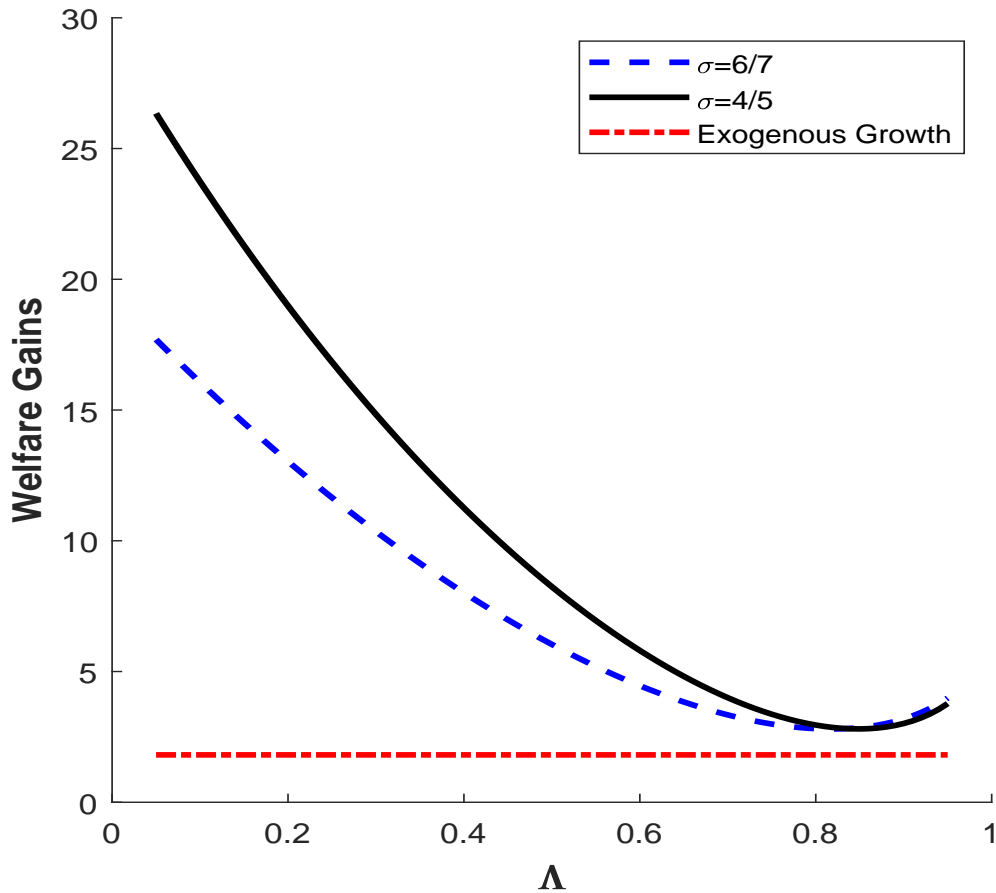
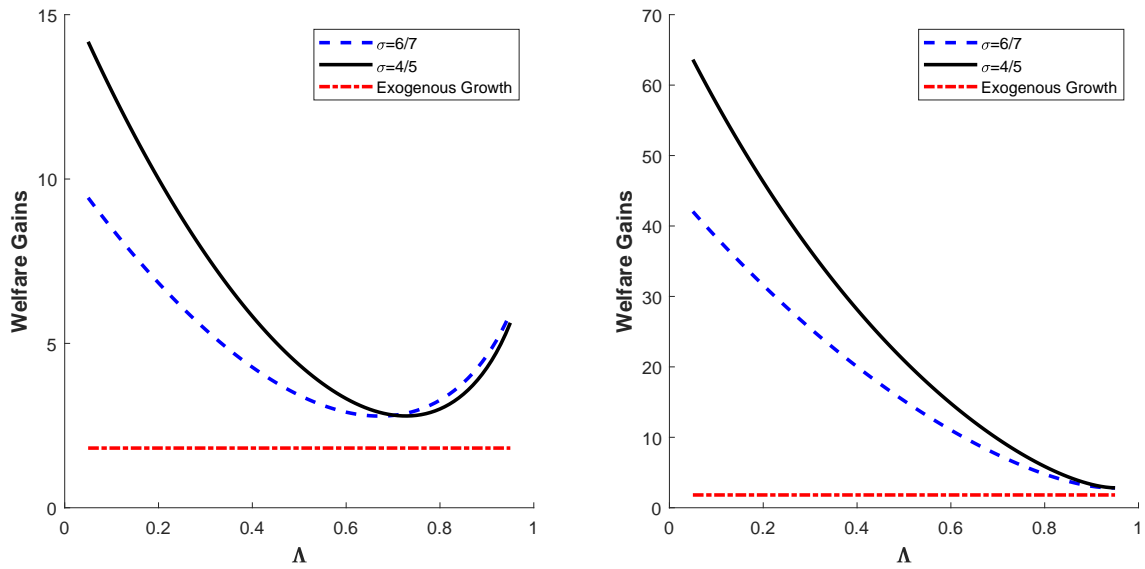


Figure 5: Welfare gains from switching to optimal policy from initial  $\tau_z = 0.4$

We now alter the profit taxes on the pre-initial BGP, to 0% in Panel (a) and 80% in Panel (b) of Figure 6. When pre-initial profit taxes are low, then the potential gains from optimal fiscal policy for low levels of  $\Lambda$  are naturally diminished, as there is not much to be improved on. When pre-initial profit taxes are high, and 80% already seems like a fairly high number, then the potential welfare gains from optimal policy are much larger; in this case, and with the most optimistic values of  $\Lambda$  and  $\sigma$ , the welfare gains rise to an astonishing 63.6%. This is in terms of public and private consumption, as an equivalent rise in private consumption only, it would be 87.5%. The welfare gains from optimal fiscal policy can thus reach remarkable heights.<sup>16</sup>

<sup>16</sup>We have used a conservative estimate for the Frisch elasticity of labour supply, 0.5. For a higher value of the elasticity, 1.0, the welfare gains almost double for values of  $\Lambda$  close to 0; for values of



(a) Low initial innovation taxes  $\tau_Z = 0.0$     (b) High initial innovation taxes  $\tau_Z = 0.8$

Figure 6: Welfare gains from optimal policy.

Finally, what are the implications of optimal policy for growth rates? In Figure 7 we show the economy growth rates as a function of  $\Lambda$ , and the two different values of  $\sigma$ , when profit taxes on the pre-initial BGP are 40%. The point where optimal growth rates are starting to be below the exogenous growth rates corresponds (roughly) to the turning point mentioned above, where the optimal profit taxes on the BGP are higher than the profit taxes on the pre-initial BGP.

$\Lambda$  close to 1 the welfare gains are essentially the same. The intuition is that, for lower values of  $\Lambda$ , optimal policy increases the incentives to innovate by reducing the taxes on innovative labour. The higher the Frisch elasticity of labour supply, the larger the labour supply effects of these incentives. When research congestion is not too strong, this leads to an increase in the growth rate. While the welfare effects of a higher Frisch elasticity may be substantial, optimal policy remains essentially the same as with a lower elasticity. The long-run growth rate under optimal policy is 2.21% with an elasticity of 1.0, as compared to 2.14% with an elasticity of 0.5.



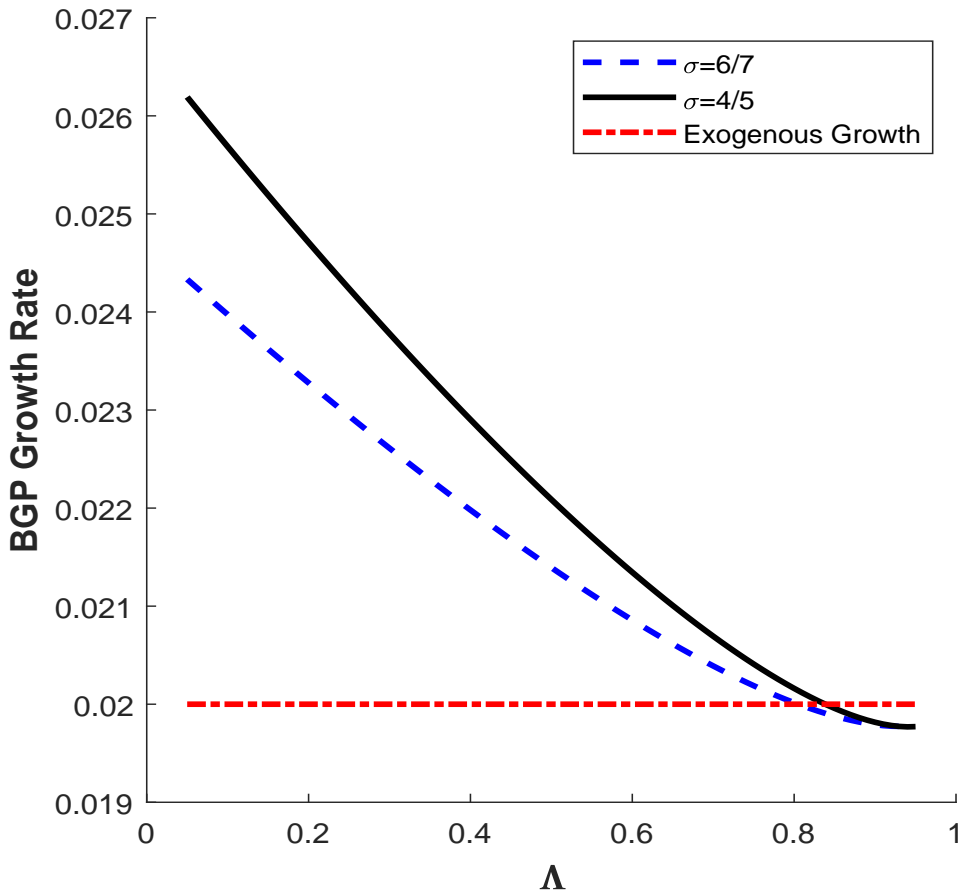


Figure 7: Yearly growth rates under switch to optimal policy from  $\tau_Z = 0.4$ .

In our benchmark case of  $\Lambda = 0.5$  and  $\sigma = 6/7$ , the growth rate increases to 2.14% per annum, up from 2.00% in the exogenous growth case. In the most optimistic case, the annual growth rate would be 2.62%, which is a more substantial increase. Not surprisingly, the welfare gains from optimal policy accrue mostly in the long run. When the pre-initial profit taxes are 80%, then the long-run growth rate can be as high as 3.17% (this result is not shown in the graphs), while the exogenous growth rate is calibrated to be the same, i.e. 2.00%.

## 6 Conclusion

We have studied optimal taxation in a variant of the Romer (1990) model. We found that capital should be subsidized and that whether innovation should be subsidized

or not depends on (i) government spending needs, (ii) the extent to which innovative activity is merely rent-seeking, and (iii) the substitutability between different varieties of intermediate goods. For a reasonable baseline specification, innovation should be subsidized rather than taxed. We also find that the welfare gains from optimal reform are much greater than in an otherwise similar exogenous growth model.

Our findings provide a new perspective on the results of Lockwood et al. (2017), who consider corrective taxation in a world consisting of several professions, each of which comes with its unique combination of non-pecuniary benefits and (negative or positive) externalities, i.e. the extent to which they are over- or underpaid relative to their contribution to society. In particular, they assume that research related activities are not fully rewarded relative to their contribution to society. They conclude that enormous gains are associated with subsidizing research professions. We are not necessarily unsympathetic to that conclusion, and it may be correct in the light of our model too. Nevertheless, our work provides a warning against jumping to the conclusion that an activity should be subsidized from the observation that it yields enormous gains for society as a whole. In our environment, innovative activity is the ultimate source of long-term growth, and so it yields enormous benefits to society. Nevertheless, because of the crowding effect that we allow for, these benefits may to a large extent be given and unresponsive to the total amount of time and effort devoted to innovation, in which case innovative activity should be taxed rather than subsidized. In any case, we view their work as complementary with ours, each providing, we hope, a useful perspective on the other.

Our work suggests a number of avenues for further research. For instance, in our model, the household consists of an innovator and a worker who share consumption. Suppose instead that households are heterogeneous; some are innovators, some are workers. How would that affect our conclusions? There would then likely be a trade-off between efficiency and equity, to the extent that innovators earn on average more than workers. Moreover, we assumed that ideas are exclusive for one period, which in our model represents 20 years. An interesting question that we did

not address is whether this patent length is optimal. What are the tradeoffs that arise from shortening or lengthening patent lives, especially in conjunction with optimal fiscal policy? Could everyone be made better off by extending or shortening patent life? Finally, how could our framework be extended to deal with a multi-country world? In such an extended framework, how would tax competition between governments affect growth?

The Romer (1990) model has been criticized for being at odds with the data, in particular that it exhibits a strong scale effect: a doubling of the population would more than double the growth rate. In our model, the scale effect is present, but much weaker, due to our assumption of a research congestion. Jones (1995) proposes a modification of the Romer model that eliminates the scale effect. The long-run growth rate is then by construction independent of policy. Nevertheless, it doesn't follow that policy analysis in such an environment is meaningless. Policy may (and probably does) matter along the transition of such a model. Since we are able to compute transitions in the context of our model, there is no major conceptual or practical difficulty facing a numerical analysis of optimal policy in a semi-endogenous growth model in the spirit of Jones (1995). We leave this for future research. Another possible avenue for future research is to quantitatively study optimal fiscal policy in a model of creative destruction such as Aghion and Howitt (1992).

## Appendix A: Decentralized equilibrium

The following system of equations has to be satisfied at each time period  $t$ :

$$\begin{aligned}
r_t &= \alpha(K_t/L_t)^{\alpha-1} \left( Z_t + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma} \\
w_t &= (1-\alpha)(K_t/L_t)^\alpha \left( Z_t + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma} \\
\Pi_{z,t} &= (1-\sigma) \sigma^{\sigma/(1-\sigma)} \frac{\left( Z_t + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma}}{Z_t + \widehat{Z}_t \sigma^{1/(1-\sigma)}} K_t^\alpha L_t^{1-\alpha} \\
R_t &= (r_t - \delta)(1 - \tau_{K,t}) \\
P_t &= \Pi_{z,t}(1 - \tau_{Z,t}) \\
A_t &= K_t + B_t \\
-u_{l,t} &= u_{c,t}(1 - \tau_{L,t})w_t \\
-u_{h,t} &= u_{c,t}P_t Z_t H_t^{-\Lambda} \eta \\
u_{c,t} &= \beta u_{c,t+1}(1 + R_{t+1}) \\
C_t &= (1 - \tau_{l,t})w_t L_t + P_t \widehat{Z}_t + (1 + R_t)A_t - A_{t+1} \\
\widehat{Z}_t &= Z_t H_t^{1-\Lambda} \eta \\
Z_{t+1} &= Z_t + \widehat{Z}_t.
\end{aligned}$$

We then have the following set of variables:

$$X_D = \left\{ r_t, w_t, \Pi_{z,t}, P_t, R_t, K_t, L_t, H_t, C_t, A_{t+1}, Z_{t+1}, \widehat{Z}_t \right\}_{t=0}^{\infty}.$$

It can be convenient to normalize the variables in such a way that they are constant on a BGP (BGP). The variables  $r_t$ ,  $R_t$ ,  $L_t$ , and  $H_t$  are all constant on a BGP without normalization; the variables  $w_t$ ,  $K_t$ ,  $C_t$ , and  $A_t$  grow at the rate  $\mathcal{G}_t := (\widehat{Z}_t/Z_t)^{(1-\sigma)/(\sigma(1-\alpha))}$ ;  $\widehat{Z}_t$  grows at the rate  $\widehat{Z}_t/Z_t$ ; finally,  $\Pi_{z,t}$  and  $P_t$  grow at the rate  $(\widehat{Z}_t/Z_t)^{(1-\sigma(2-\alpha))/(\sigma(1-\alpha))}$ . The resulting system of equations with appropriately scaled variables, where the intertemporal elasticity of substitution is  $1/\zeta$ , is

then

$$\begin{aligned}
r_t &= \alpha(K_t/L_t)^{\alpha-1} \left(1 + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)}\right)^{(1-\sigma)/\sigma}, \\
w_t &= (1-\alpha)(K_t/L_t)^\alpha \left(1 + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)}\right)^{(1-\sigma)/\sigma}, \\
\Pi_{z,t} &= (1-\sigma)\sigma^{\sigma/(1-\sigma)} \frac{\left(1 + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)}\right)^{(1-\sigma)/\sigma}}{1 + \widehat{Z}_t \sigma^{1/(1-\sigma)}} K_t^\alpha L_t^{1-\alpha}, \\
R_t &= (r_t - \delta)(1 - \tau_{K,t}), \\
P_t &= \Pi_{z,t}(1 - \tau_{Z,t}), \\
A_t &= K_t + B_t, \\
-u_{l,t} &= u_{c,t}(1 - \tau_{L,t})w_t, \\
-u_{h,t} &= u_{c,t}P_t H_t^{-\Lambda} \eta, \\
u_{c,t}(1 + \mathcal{G}_t)^\zeta &= \beta u_{c,t+1}(1 + R_{t+1}), \\
C_t &= (1 - \tau_{l,t})w_t L_t + P_t \widehat{Z}_t + (1 + R_t)A_t - A_{t+1}(1 + g_t), \\
\text{and} \\
\widehat{Z}_t &= H_t^{1-\Lambda} \eta.
\end{aligned}$$

## Appendix B: Derivations

The Lagrangian for a monopolist is

$$\mathcal{L}_{z,t} = Q_{z,t}^\sigma Y_t^{1-\sigma} - r_t K_{z,t} - w_t L_{z,t} + MC_{z,t} \left[ K_{z,t}^\alpha L_{z,t}^{1-\alpha} - Q_{z,t} \right]. \quad (30)$$

As is well known, the Lagrange multiplier is equal to marginal cost, hence the notation. The first-order conditions are

$$r_t = MC_{z,t} \alpha (L_{z,t}/K_{z,t})^{1-\alpha} \quad (31)$$

$$w_t = MC_{z,t} (1-\alpha) (K_{z,t}/L_{z,t})^\alpha \quad (32)$$

$$MC_{z,t} = \sigma Q_{z,t}^{\sigma-1} Y_t^{1-\sigma}. \quad (33)$$

Dividing Equation (31) by (32), solving for  $K_{z,t}/L_{z,t}$ , and re-inserting into (31) we obtain

$$MC_{z,t} = \left( \frac{r_t}{\alpha} \right)^\alpha \left( \frac{w_t}{1-\alpha} \right)^{1-\alpha}.$$

We can now write profits as

$$\Pi_{z,t} = (1-\sigma) \left( \frac{\sigma}{MC} \right)^{\sigma/(1-\sigma)} Y_t.$$

Final output is given by

$$Y_t = \left( Z_t Q_{z,t}^C + \widehat{Z}_t Q_{z,t}^M \right)^{1/\sigma},$$

so using  $Q_{z,t}^C = \left( \frac{1}{MC_{z,t}} \right)^{1/(1-\sigma)} Y$  and  $Q_{z,t}^M = \left( \frac{\sigma}{MC_{z,t}} \right)^{1/(1-\sigma)} Y$  implies that

$$MC_{z,t} = \left( Z_t + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma}.$$

We now use the fact that

$$Y_t = r_t K_t + w_t L_t + \Pi_t \widehat{Z}_t$$

together with the results above to obtain

$$Y_t = \frac{\left(Z_t + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)}\right)^{1/\sigma}}{Z_t + \widehat{Z}_t \sigma^{1/(1-\sigma)}} K_t^\alpha L_t^{1-\alpha}.$$

## Appendix C: Proofs

**Proposition 1** If the economy converges to a non-degenerate BGP, then the modified golden rule holds. The capital share, net of taxes, of total income is given by

$$\frac{(1 - \tau_{K,t})(r_t - \delta)K_t}{Y_t} = \frac{\partial Y_t}{\partial K_t} \cdot \frac{K_t}{Y_t} - \delta \frac{K_t}{Y_t}.$$

**Proof.** Public consumption  $G_t$  grows at a constant, non-negative rate on a non-degenerate BGP, hence the marginal utility of public consumption evolves according to

$$v_{G,t+1} = v_{G,t}/(1 + \mathcal{G}_t)^\zeta, \quad (34)$$

where  $\mathcal{G}_t$  is the rate of growth of output (as well as public and private consumption and capital), and the intertemporal elasticity of substitution is  $1/\zeta$ . The first-order condition with respect to  $G_t$  is

$$\beta^t v_{G,t} = \psi_t, \quad (35)$$

so we then have

$$\psi_{t+1} = \beta \psi_t / (1 + \mathcal{G}_t)^\zeta. \quad (36)$$

From the household's intertemporal optimality condition, Equation (14), we know that

$$\beta(1 + R_{t+1}) = (1 + \mathcal{G}_t)^\zeta. \quad (37)$$

The first-order condition with respect to  $B_{t+1}$  is

$$\psi_t = \psi_{t+1}(1 + R_{t+1}) - \omega_{t+1}, \quad (38)$$

which leads us to conclude that  $\omega_{t+1} = 0$  on a non-degenerate BGP. We also know that  $\phi_t = 0$  on such a BGP, otherwise  $R_t = 0$  and hence  $1 > \beta = (1 + \mathcal{G}_t)^\zeta > 1$ , a contradiction; and clearly,  $\varphi_t = 0$ , otherwise  $\mathcal{G}_t = 0$ . The first-order condition with



respect to  $K_t$  is

$$\begin{aligned}
& \psi_t \tau_{K,t} (r_t - \delta) + \psi_t \tau_{K,t} \frac{\partial r(K_t, L_t, Z_t, \hat{Z}_t)}{\partial K_t} K_t + \psi_t \tau_{L,t} \frac{\partial w(K_t, L_t, Z_t, \hat{Z}_t)}{\partial K_t} L_t + \\
& \psi_t \tau_{Z,t} \frac{\partial \Pi(K_t, L_t, Z_t, \hat{Z}_t)}{\partial K_t} \hat{Z}_t + \theta_t (1 - \tau_{L,t}) \frac{\partial w(K_t, L_t, Z_t, \hat{Z}_t)}{\partial K_t} L_t + \\
& \mu_t (1 - \tau_{L,t}) \frac{\partial w(K_t, L_t, Z_t, \hat{Z}_t)}{\partial K_t} u_{C,t} + \gamma_t \frac{\partial r(K_t, L_t, Z_t, \hat{Z}_t)}{\partial K_t} (1 - \tau_{K,t}) + \\
& \gamma_t \frac{\partial \Pi(K_t, L_t, Z_t, \hat{Z}_t)}{\partial K_t} (1 - \tau_{Z,t}) = \omega_t. \quad (39)
\end{aligned}$$

From here on, we suppress the arguments of  $r_t$ ,  $w_t$ , and  $\Pi_t$  for notational convenience. Now we can combine this with the first-order conditions with respect to  $\tau_{K,t}$ ,  $\tau_{L,t}$ , and  $\tau_{Z,t}$  (and using the fact that  $\phi_t = \varphi_t = 0$  on a non-degenerate BGP),

$$\psi_t (r_t - \delta) K_t = \gamma_t (r_t - \delta) \quad (40)$$

$$\psi_t L_t - \theta_t L_t = \mu_t u_{C,t} \quad (41)$$

$$\psi_t \hat{Z}_t = \gamma_t, \quad (42)$$

to arrive at

$$\psi_t \left[ \tau_{K,t} (r_t - \delta) + \frac{\partial r}{\partial K_t} K_t + \frac{\partial w}{\partial K_t} L_t + \frac{\partial \Pi}{\partial K_t} \hat{Z}_t \right] = 0. \quad (43)$$

We can now use the identity that  $Y_t = r_t K_t + w_t L_t + \Pi_t \hat{Z}_t$  and hence

$$\frac{\partial Y}{\partial K_t} - r_t = \frac{\partial r}{\partial K_t} K_t + \frac{\partial w}{\partial K_t} L_t + \frac{\partial \Pi}{\partial K_t} \hat{Z}_t \quad (44)$$

to conclude that

$$(1 - \tau_{K,t})(r_t - \delta) = \frac{\partial Y}{\partial K_t} - \delta. \quad (45)$$

In words, the after-tax return on capital is equal to the marginal product of capital. Together with the household's intertemporal optimality condition, Equation (14), this implies that the modified golden rule holds:

$$u_{c,t} = \beta u_{c,t+1} \left( 1 + \frac{\partial Y}{\partial K_t} - \delta \right). \quad (46)$$

Multiplying Equation (45) by  $K_t/Y_t$  yields Equation (26). ■

**Corollary 1** Capital taxes on a non-degenerate BGP are negative.

**Proof.** In order to equate the private returns to capital to the marginal product of capital, a corrective negative tax is necessary, since  $\frac{\partial Y}{\partial K_t} \neq r_t$ . The tax is given by

$$\begin{aligned}
\tau_{K,t}(r_t - \delta) &= r_t - \frac{\partial Y}{\partial K_t} \\
&= \alpha \frac{K_t^{\alpha-1}}{L_t} \left( Z_t + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma} - \alpha \frac{K_t^{\alpha-1}}{L_t} \frac{\left( Z_t + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{1/\sigma}}{Z_t + \widehat{Z}_t \sigma^{1/(1-\sigma)}} \\
&= \alpha \frac{K_t^{\alpha-1}}{L_t} \frac{\left( Z_t + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma}}{Z_t + \widehat{Z}_t \sigma^{1/(1-\sigma)}} \left[ Z_t + \widehat{Z}_t \sigma^{1/(1-\sigma)} - Z_t - \widehat{Z}_t \sigma^{\sigma/(1-\sigma)} \right] \\
\tau_{K,t}(r_t - \delta) &= \alpha \frac{K_t^{\alpha-1}}{L_t} \frac{\left( Z_t + \widehat{Z}_t \sigma^{\sigma/(1-\sigma)} \right)^{(1-\sigma)/\sigma}}{Z_t + \widehat{Z}_t \sigma^{1/(1-\sigma)}} \widehat{Z}_t \sigma^{1/(1-\sigma)} (1 - 1/\sigma). \tag{47}
\end{aligned}$$

Since  $\sigma < 1$ , it follows that  $\tau_{K,t} < 0$ . ■

**Proposition 2** If the economy converges to a non-degenerate BGP and if cross-derivatives of the period utility function are identically equal to zero, the labour share, net of taxes, of total income is given by

$$\frac{(1 - \tau_{L,t})w_t L_t}{Y_t} = \frac{\frac{\partial Y_t}{\partial L_t} \cdot \frac{L_t}{Y_t}}{1/\tilde{\psi}_t + \left(1 - \tilde{\theta}_t/\tilde{\psi}_t\right) (1 + \epsilon_{L,t})}.$$

**Proof.** When cross-derivatives of the period utility function are identically equal to zero, the first-order condition with respect to labour at time t is

$$\begin{aligned}
\beta^t u_{L,t} + \psi_t \left[ \tau_{K,t} \frac{\partial r}{\partial L_t} K_t + \tau_{L,t} \frac{\partial w}{\partial L_t} L_t + \tau_{Z,t} \frac{\partial \pi}{\partial L_t} \widehat{Z}_t \right] + \psi_t \tau_{L,t} w_t + \theta_t (1 - \tau_{L,t}) w_t + \tag{48} \\
\theta_t (1 - \tau_{L,t}) \frac{\partial w}{\partial L_t} L_t + \mu_t (1 - \tau_{L,t}) \frac{\partial w}{\partial L_t} u_{C,t} + \mu_t u_{LL,t} + \gamma_t (1 - \tau_{K,t}) \frac{\partial r}{\partial L_t} + \Upsilon_t (1 - \tau_{Z,t}) \frac{\partial \pi}{\partial L_t} = 0.
\end{aligned}$$

Making use of the optimality conditions with respect to the tax instruments, Equa-

tions (40) to (42), and assuming a non-degenerate BGP, Equation (48) becomes

$$\psi_t \left[ \frac{\partial r}{\partial L_t} K_t + \frac{\partial w}{\partial L_t} L_t + \frac{\partial \pi}{\partial L_t} \widehat{Z}_t \right] + \psi_t \tau_{L,t} w_t + \theta_t (1 - \tau_{L,t}) w_t = -\beta^t u_{L,t} - \mu_t u_{LL,t}. \quad (49)$$

We use again the identity that  $Y_t = r_t K_t + w_t L_t + \Pi_t \widehat{Z}_t$  and thus

$$\frac{\partial Y}{\partial L_t} - w_t = \frac{\partial r}{\partial L_t} K_t + \frac{\partial w}{\partial L_t} L_t + \frac{\partial \Pi}{\partial L_t} \widehat{Z}_t \quad (50)$$

to transform Equation (49) into

$$\psi_t \frac{\partial Y}{\partial L_t} = -\beta^t u_{L,t} - \mu_t u_{LL,t} + (\psi_t - \theta_t) (1 - \tau_{L,t}) w_t \quad (51)$$

Since  $\mu_t = (\psi_t - \theta_t) L_t / u_{C,t}$  and  $u_{L,t} = -(1 - \tau_{L,t}) w_t u_{C,t}$ , this becomes

$$\psi_t \frac{\partial Y}{\partial L_t} = (1 - \tau_{L,t}) w_t \left[ \beta^t u_{C,t} + (\psi_t - \theta_t) (1 + u_{LL,t} L_t / u_{L,t}) \right]. \quad (52)$$

Dividing everything by  $\beta^t u_{C,t}$  and using the definitions for  $\tilde{\psi}_t = \frac{\psi_t}{\beta^t u_{C,t}}$ ,  $\tilde{\theta}_t = \frac{\theta_t}{\beta^t u_{C,t}}$ , and  $\epsilon_{L,t} = \frac{u_{LL,t} L_t}{u_{L,t}}$ , we obtain

$$\tilde{\psi}_t \frac{\partial Y}{\partial L_t} = (1 - \tau_{L,t}) w_t \left[ 1 + (\tilde{\psi}_t - \tilde{\theta}_t) (1 + \epsilon_{L,t}) \right]. \quad (53)$$

Dividing both sides by  $Y_t$  and  $\tilde{\psi}_t$ , multiplying by  $L_t$ , and doing simple algebraic manipulations completes the proof:

$$\frac{(1 - \tau_{L,t}) w_t L_t}{Y_t} = \frac{\frac{\partial Y_t}{\partial L_t} \cdot \frac{L_t}{Y_t}}{1 / \tilde{\psi}_t + (1 - \tilde{\theta}_t / \tilde{\psi}_t) (1 + \epsilon_{L,t})}. \quad \blacksquare$$

**Lemma 1** If the economy converges to a non-degenerate BGP and if cross-derivatives of the period utility function are identically equal to zero, as long as lump-sum taxes are unavailable, then  $\psi_t > \theta_t$ , and

$$1 / \tilde{\psi}_t + (1 - \tilde{\theta}_t / \tilde{\psi}_t) > 1.$$

**Proof.** It is natural to think that  $\psi_t - \theta_t > 0$ , and it is easy to show that this indeed must be true in the absence of lump-sum taxes. If we introduced lump-sum taxes  $\vartheta_t$ , they would show up in the government and household budget constraint; if we furthermore introduced a constraint that limits these taxes,  $\vartheta_t \leq \bar{\vartheta}$ , with a Lagrange multiplier  $\varsigma_t$ , then the first-order condition with respect to  $\vartheta_t$  is  $\psi_t - \theta_t - \varsigma_t = 0$ . Hence, as long as the constraint is binding,  $\psi_t - \theta_t > 0$ .

Multiplying both sides of the inequality in the Lemma by  $\tilde{\psi}_t > 0$  yields  $\tilde{\theta}_t < 1$ . Using the definition of  $\tilde{\theta}_t$ , this is equivalent to  $\theta_t < \beta^t u_{C,t}$ . The first-order condition with respect to consumption at time  $t$ , assuming that the cross-derivatives of the period utility function are identically equal to zero, is

$$\beta^t u_{C,t} - \theta_t + u_{CC,t} \left[ \mu_t (1 - \tau_{L,t}) w_t + \lambda_t \eta P_t Z_t H_t^\Lambda + \zeta_{t-1} \beta (1 + R_t) - \zeta_t \right] = 0. \quad (54)$$

On a BGP,  $\zeta_t = \beta \zeta_{t-1} (1 + \mathcal{G}_t)$ , so we can rewrite Equation (54) as

$$\theta_t = \beta^t u_{C,t} + u_{CC,t} \left[ \mu_t (1 - \tau_{L,t}) w_t + \lambda_t \eta P_t Z_t H_t^\Lambda + \zeta_{t-1} \beta (R_t - \mathcal{G}_t) \right]. \quad (55)$$

Since  $u_{CC,t} < 0$  and  $R_t > \mathcal{G}_t$  on a non-degenerate BGP (and all prices and quantities are naturally positive), if the Lagrange multipliers  $\mu_t$ ,  $\lambda_t$ , and  $\zeta_{t-1}$  are all non-negative, then the inequality is proven. If  $R_t$  were smaller or equal to  $\mathcal{G}_t$ , then the household could increase consumption by decreasing assets, i.e. it would violate the transversality condition. Making use of the optimality condition for labour taxes, Equation (41), it follows that

$$\mu_t = (\psi_t - \theta_t) L_t / u_{C,t} > 0. \quad (56)$$

The first-order conditions with respect to profit taxes, Equation (42), and with respect to the patent price,

$$\theta_t (\Delta Z)_t + \lambda_t \eta u_{C,t} Z_t H_t^{-\Lambda} - \gamma_t = 0, \quad (57)$$

imply together with the production of patents, Equation (8), that

$$\lambda_t = (\psi_t - \theta_t)H_t/u_{C,t} > 0. \quad (58)$$

Finally, the first-order condition with respect to the net interest rate  $R_t$  is

$$\psi_t B_t + \theta_t A_t + \zeta_{t-1} \beta u_{C,t} - \gamma_t = 0. \quad (59)$$

Combining this with the optimality condition for capital taxes, Equation (40), and making use of the asset-market clearing condition, Equation (16), we get

$$\beta \zeta_{t-1} = (\psi_t - \theta_t)A_t/u_{C,t} > 0. \quad (60)$$

■

**Proposition 3** If the economy converges to a non-degenerate BGP and if cross-derivatives of the period utility function are identically equal to zero, the profit share, net of taxes, of total income is given by

$$\frac{(1 - \tau_{Z,t})\Pi_t \hat{Z}_t}{Y_t} = (1 - \Lambda) \cdot \frac{\frac{\partial Y_t/Y_t}{\partial \hat{Z}_t/\hat{Z}_t} + \frac{\beta \hat{Z}_t/Z_t}{(1-\beta)(1+\hat{Z}_t/Z_t)} \left( \frac{\partial Y_t}{\partial Z_t} \cdot \frac{Z_t}{Y_t} + \frac{\partial Y_t}{\partial \hat{Z}_t} \cdot \frac{\hat{Z}_t}{Y_t} \right)}{1/\tilde{\psi}_t + \left(1 - \tilde{\theta}_t/\tilde{\psi}_t\right) (1 + \epsilon_{H,t})}.$$

**Proof.** We start with the first-order condition for  $Z_{t+1}$  at time  $t - 1$ :

$$\begin{aligned} & \psi_t \tau_{K,t} \frac{\partial r}{\partial Z_t} K_t + \psi_t \tau_{L,t} \frac{\partial w}{\partial Z_t} L_t + \psi_t \tau_{Z,t} \frac{\partial \Pi}{\partial Z_t} \hat{Z}_t + \\ & \theta_t (1 - \tau_{L,t}) \frac{\partial w}{\partial Z_t} L_t + \mu_t (1 - \tau_{L,t}) \frac{\partial w}{\partial Z_t} u_{C,t} + \\ & \gamma_t (1 - \tau_{K,t}) \frac{\partial r}{\partial Z_t} + \Upsilon_t (1 - \tau_{Z,t}) \frac{\partial \Pi}{\partial Z_t} + \\ & \lambda_t \eta u_{C,t} P_t H_t^{-\Lambda} + \Omega_t \eta H_t^{1-\Lambda} + \kappa_t - \kappa_{t-1} = 0. \end{aligned} \quad (61)$$

As above, we make use of the three first-order conditions with respect to the tax

instruments, Equations (40) to (42), and assuming a non-degenerate BGP:

$$\begin{aligned} & \psi_t \frac{\partial r}{\partial Z_t} K_t + \psi_t \frac{\partial w}{\partial Z_t} L_t + \psi_t \frac{\partial \Pi}{\partial Z_t} \hat{Z}_t + \\ & \lambda_t \eta u_{C,t} P_t H_t^{-\Lambda} + \Omega_t \eta H_t^{1-\Lambda} + \kappa_t - \kappa_{t-1} = 0. \end{aligned} \quad (62)$$

In order to substitute out  $\lambda_t$ , we employ the first-order condition with respect to profit taxes, Equation (42), and with respect to the patent price, Equation (57), to arrive at

$$\begin{aligned} & \psi_t \frac{\partial r}{\partial Z_t} K_t + \psi_t \frac{\partial w}{\partial Z_t} L_t + \psi_t \frac{\partial \Pi}{\partial Z_t} \hat{Z}_t + \\ & (\psi_t - \theta_t) P_t \frac{\hat{Z}_t}{Z_t} + \Omega_t \eta H_t^{1-\Lambda} + \kappa_t - \kappa_{t-1} = 0. \end{aligned} \quad (63)$$

We make use of the production of patents, Equation (8), and the marginal product of  $Z_t$ ,

$$\frac{\partial Y}{\partial Z_t} = \frac{\partial r}{\partial Z_t} K_t + \frac{\partial w}{\partial Z_t} L_t + \frac{\partial \Pi}{\partial Z_t} \hat{Z}_t, \quad (64)$$

which implies that

$$\psi_t \frac{\partial Y}{\partial Z_t} + (\psi_t - \theta_t) P_t \frac{\hat{Z}_t}{Z_t} + \Omega_t \frac{\hat{Z}_t}{Z_t} + \kappa_t - \kappa_{t-1} = 0. \quad (65)$$

A non-degenerate BGP where the cross-derivatives of the period utility function are identically equal to zero implies an elasticity of intertemporal substitution of one. It follows that  $\kappa_{t-1} = \kappa_t \left(1 + \frac{\hat{Z}_t}{Z_t}\right) / \beta$ , and so

$$\psi_t \frac{\partial Y}{\partial Z_t} + (\psi_t - \theta_t) P_t \frac{\hat{Z}_t}{Z_t} + \Omega_t \frac{\hat{Z}_t}{Z_t} = \kappa_t \left(1 - \beta + \frac{\hat{Z}_t}{Z_t}\right) / \beta. \quad (66)$$

In order to substitute out the Lagrange multiplier  $\kappa_t$ , we have to turn to the first-

order condition for  $\widehat{Z}_t$ :

$$\begin{aligned}
& \psi_t \tau_{K,t} \frac{\partial r}{\widehat{Z}_t} K_t + \psi_t \tau_{L,t} \frac{\partial w}{\widehat{Z}_t} L_t + \psi_t \tau_{Z,t} \frac{\partial \Pi}{\widehat{Z}_t} \widehat{Z}_t + \\
& \theta_t (1 - \tau_{L,t}) \frac{\partial w}{\widehat{Z}_t} L_t + \mu_t (1 - \tau_{L,t}) \frac{\partial w}{\widehat{Z}_t} u_{C,t} + \\
& \gamma_t \frac{\partial r}{\widehat{Z}_t} (1 - \tau_{K,t}) + \Upsilon_t \frac{\partial \Pi}{\widehat{Z}_t} (1 - \tau_{Z,t}) + \\
& \psi_t \tau_{Z,t} \Pi_t + \theta_t P_t + \kappa_t - \Omega_t = 0.
\end{aligned} \tag{67}$$

Making use again of Equations (40) to (42), this becomes:

$$\begin{aligned}
& \psi_t \frac{\partial r}{\widehat{Z}_t} K_t + \psi_t \frac{\partial w}{\widehat{Z}_t} L_t + \psi_t \frac{\partial \Pi}{\widehat{Z}_t} \widehat{Z}_t + \\
& \psi_t \tau_{Z,t} \Pi_t + \theta_t P_t + \kappa_t - \Omega_t = 0.
\end{aligned} \tag{68}$$

The marginal product of  $\widehat{Z}_t$  is

$$\frac{\partial Y}{\partial \widehat{Z}_t} = \frac{\partial r}{\partial \widehat{Z}_t} K_t + \frac{\partial w}{\partial \widehat{Z}_t} L_t + \frac{\partial \Pi}{\partial \widehat{Z}_t} \widehat{Z}_t + \Pi_t, \tag{69}$$

and we thus have

$$\kappa_t = -\psi_t \frac{\partial Y}{\partial \widehat{Z}_t} + (\psi_t - \theta_t) P_t + \Omega_t. \tag{70}$$

In order to substitute out the Lagrange multiplier  $\Omega_t$ , we employ the first-order condition with respect to  $H_t$  (noting that we are considering the case where cross-derivatives of the period utility function are identically equal to zero):

$$\beta^t u_{H,t} - \lambda_t \Lambda \eta u_{C,t} P_t Z_t H_t^{-\Lambda-1} + \lambda_t u_{HH,t} + \Omega_t \eta (1 - \Lambda) Z_t H_t^{-\Lambda} = 0. \tag{71}$$

Using Equations (42), (8), and (57) this becomes

$$\Omega_t = -\frac{\beta^t u_{H,t} H_t}{\widehat{Z}_t (1 - \Lambda)} + \frac{\Lambda}{1 - \Lambda} (\psi_t - \theta_t) P_t - \frac{1}{1 - \Lambda} (\psi_t - \theta_t) \frac{u_{HH,t} H_t^2}{u_{C,t} \widehat{Z}_t}. \tag{72}$$

Combining Equations (66), (70), and (72), we obtain

$$\begin{aligned} \frac{\psi_t(1-\Lambda)}{(1-\beta)(1+\widehat{Z}_t/Z_t)} \left[ \frac{\partial Y}{\partial Z_t/Z_t} \beta \widehat{Z}_t/Z_t + \frac{\partial Y}{\partial \widehat{Z}_t/\widehat{Z}_t} \left(1-\beta + \widehat{Z}_t/Z_t\right) \right] \\ = -\beta^t u_{H,t} H_t + (\psi_t - \theta_t) \left[ P_t \widehat{Z}_t - \frac{u_{HH,t} H_t^2}{u_{C,t}} \right]. \end{aligned} \quad (73)$$

Now we use the fact that  $u_{H,t} = -u_{C,t} P_t \widehat{Z}_t / H_t$  to rewrite this as

$$\begin{aligned} \frac{\psi_t(1-\Lambda)}{(1-\beta)(1+\widehat{Z}_t/Z_t)} \left[ \frac{\partial Y}{\partial Z_t/Z_t} \beta \widehat{Z}_t/Z_t + \frac{\partial Y}{\partial \widehat{Z}_t/\widehat{Z}_t} \left(1-\beta + \widehat{Z}_t/Z_t\right) \right] \\ = \beta^t u_{C,t} P_t \widehat{Z}_t + (\psi_t - \theta_t) \left[ P_t \widehat{Z}_t + \frac{u_{HH,t} H_t}{u_{H,t}} P_t \widehat{Z}_t \right]. \end{aligned} \quad (74)$$

Dividing everything by  $\beta^t u_{C,t}$  and using the definitions for  $\tilde{\psi}_t = \frac{\psi_t}{\beta^t u_{C,t}}$ ,  $\tilde{\theta}_t = \frac{\theta_t}{\beta^t u_{C,t}}$ , and  $\epsilon_{H,t} = \frac{u_{HH,t} H_t}{u_{H,t}}$ , we obtain

$$\begin{aligned} \frac{\tilde{\psi}_t(1-\Lambda)}{(1-\beta)(1+\widehat{Z}_t/Z_t)} \left[ \frac{\partial Y}{\partial Z_t/Z_t} \beta \widehat{Z}_t/Z_t + \frac{\partial Y}{\partial \widehat{Z}_t/\widehat{Z}_t} \left(1-\beta + \widehat{Z}_t/Z_t\right) \right] \\ = P_t \widehat{Z}_t \left[ 1 + (\tilde{\psi}_t - \tilde{\theta}_t)(1 + \epsilon_{H,t}) \right]. \end{aligned} \quad (75)$$

Dividing both sides by  $Y_t$  and  $\tilde{\psi}_t$  and performing simple algebraic manipulations, while noting that  $P_t = \pi_t(1 - \tau_{z,t})$ , completes the proof:

$$\frac{(1 - \tau_{z,t}) \Pi_t \widehat{Z}_t}{Y_t} = (1 - \Lambda) \cdot \frac{\frac{\partial Y_t/Y_t}{\partial \widehat{Z}_t/\widehat{Z}_t} + \frac{\beta \widehat{Z}_t/Z_t}{(1-\beta)(1+\widehat{Z}_t/Z_t)} \left( \frac{\partial Y_t}{\partial Z_t} \cdot \frac{Z_t}{Y_t} + \frac{\partial Y_t}{\partial \widehat{Z}_t} \cdot \frac{\widehat{Z}_t}{Y_t} \right)}{\left[ 1/\tilde{\psi}_t + \left(1 - \tilde{\theta}_t/\tilde{\psi}_t\right) (1 + \epsilon_{H,t}) \right]}.$$

■

## Appendix D: Non-zero cross derivatives

The condition for optimal capital taxes is not affected by non-zero cross derivatives in the utility function. The conditions for optimal labour and profit taxes do change, however, and we state them here. The proofs proceed along the lines of the proofs above, so we omit them.



**Proposition 2'** If the economy converges to a non-degenerate BGP the labour share, net of taxes, of total income is given by

$$\frac{(1 - \tau_{L,t})w_t L_t}{Y_t} = \frac{\frac{\partial Y_t}{\partial L_t} \cdot \frac{L_t}{Y_t}}{1/\tilde{\Psi}_t + \left(1 - \tilde{\theta}_t/\tilde{\Psi}_t\right) \left(1 + \epsilon_{L,t} - \epsilon_{CL,t} + \epsilon_{LH,t} + \epsilon_{LC,t} \left[ \frac{A_t(R_t - \mathcal{G}_t)}{C_t} + \frac{P_t \hat{Z}_t}{C_t} \right] \right)}, \quad (76)$$

where  $\epsilon_{CL,t} := u_{CL,t} \cdot \frac{L_t}{u_{C,t}}$  is the cross-elasticity of the marginal utility of consumption and working more, while  $\epsilon_{LC,t} := u_{LC,t} \cdot \frac{C_t}{u_{L,t}}$  is the cross-elasticity of the marginal disutility from working and consuming more, and  $\epsilon_{LH,t} := u_{LH,t} \cdot \frac{H_t}{u_{L,t}}$  is the cross-elasticity of the marginal disutility from working and innovating more.

**Proposition 3'** If the economy converges to a non-degenerate BGP the profit share, net of taxes, of total income is given by

$$\frac{(1 - \tau_{Z,t})\Pi_t \hat{Z}_t}{Y_t} = \frac{(1 - \Lambda) \left[ \frac{\partial Y_t / Y_t}{\partial \hat{Z}_t / \hat{Z}_t} + \frac{\tilde{\beta}_t \hat{Z}_t / Z_t}{(1 - \tilde{\beta}_t)(1 + \hat{Z}_t / Z_t)} \left( \frac{\partial Y_t}{\partial Z_t} \cdot \frac{Z_t}{Y_t} + \frac{\partial Y_t}{\partial \hat{Z}_t} \cdot \frac{\hat{Z}_t}{Y_t} \right) \right]}{1/\tilde{\Psi}_t + \left(1 - \tilde{\theta}_t/\tilde{\Psi}_t\right) \left(1 + \epsilon_{H,t} - \epsilon_{CH,t} + \epsilon_{HL,t} + \epsilon_{HC,t} \left[ \frac{A_t(R_t - \mathcal{G}_t)}{C_t} + \frac{(1 - \tau_{L,t})w_t L_t}{C_t} \right] \right)}, \quad (77)$$

where  $\tilde{\beta}_t = \beta(1 + \mathcal{G}_t)^{1-\varsigma}$ ,  $\epsilon_{CH,t} := u_{CH,t} \cdot \frac{H_t}{u_{C,t}}$  is the cross-elasticity of the marginal utility of consumption and innovating more, while  $\epsilon_{HC,t} := u_{HC,t} \cdot \frac{C_t}{u_{H,t}}$  is the cross-elasticity of the marginal disutility from innovating and consuming more, and  $\epsilon_{HL,t} := u_{HL,t} \cdot \frac{L_t}{u_{H,t}}$  is the cross-elasticity of the marginal disutility from innovating and working more.

Note that the discount factor is different here, which stems from the fact that the Lagrange multiplier dynamics over time are somewhat different when the intertemporal elasticity of substitution is not equal to one. In particular, the growth rate of  $\kappa_t$  is then  $\beta(1 + \mathcal{G}_t)^{1-\varsigma}/(1 + \hat{Z}_t/Z_t)$ .

## Appendix E: Exogenous Growth Model

There are no intermediate goods and the production function of the final good is given by

$$Y_t = K_t^\alpha (Z_t L_t)^{1-\alpha},$$

where  $Z_t$  is the exogenously specified labour productivity in period  $t$ . Technological growth is constant, so we have  $Z_{t+1} = (1 + GR)Z_t$ , where we set  $GR$  so that the annual growth rate is 2% as in our baseline calibration.

The Lagrangian is

$$\begin{aligned} \mathcal{L}^E = & \sum_{t=0}^{\infty} \beta^t \{u(C_t, L_t) + v(G_t)\} \\ & + \psi_t [\tau_{k,t}(r_t - \delta)K_t + \tau_{l,t}w_t L_t + B_{t+1} - G_t - B_t(1 + R_t)] \\ & + \theta_t [(1 - \tau_{l,t})w_t L_t + (1 + R_t)A_t - A_{t+1} - C_t] \\ & + \mu_t [(1 - \tau_{l,t})w_t u_{c,t} + u_{l,t}] \\ & + \zeta_t [\beta u_{c,t+1}(1 + R_{t+1}) - u_{c,t}] \\ & + \gamma_t [(r_t - \delta)(1 - \tau_{k,t}) - R_t] \\ & + \omega_t [A_t - K_t - B_t] \\ & + \phi_t [1 - \tau_{k,t}], \end{aligned} \tag{78}$$

with the following set of choice variables:

$$\tilde{X}^E = \{C_t, L_t, K_t, A_{t+1}, G_t, B_{t+1}, \tau_{k,t+1}, \tau_{l,t}, R_t\}_{t=0}^{\infty}. \tag{79}$$

We also have to keep in mind that  $r_t$  and  $w_t$  are functions of  $K_t$  and  $L_t$ :

$$\begin{aligned} r_t &= \alpha Z_t^{1-\alpha} (K_t/L_t)^{\alpha-1} \\ w_t &= (1 - \alpha) Z_t^{1-\alpha} (K_t/L_t)^{\alpha}. \end{aligned}$$

## References

- Aghion, P., U. Akcigit, and J. Fernández-Villaverde (2013, May). Optimal capital versus labor taxation with innovation-led growth. Working Paper 19086, National Bureau of Economic Research.
- Aghion, P. and P. Howitt (1992, March). A model of growth through creative destruction. *Econometrica* 60(2), 323–351.
- Aiyagari, S. R. and E. R. McGrattan (1998). The optimum quantity of debt. *Journal of Monetary Economics* 42(3), 447–469.
- Anderson, J. E. and E. van Wincoop (2004, September). Trade costs. *Journal of Economic Literature* 42(3), 691–751.
- Annicchiarico, B., V. Antonaroli, and A. Pelloni (2020, June). Optimal Factor Taxation in A Scale Free Model of Vertical Innovation. Working Paper series 20-20, Rimini Centre for Economic Analysis.
- 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), 3–17.
- Cass, D. (1965). Optimum growth in an aggregative model of capital accumulation. *Review of Economic Studies* 32, 233–240.
- Chamley, C. (1986). Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica* 54, 607–22.
- Conesa, J. C., S. Kitao, and D. Krueger (2009, March). Taxing capital? Not a bad idea after all! *American Economic Review* 99(1), 25–48.
- Correia, I. H. (1996, April). Should capital income be taxed in the steady state? *Journal of Public Economics* 60(1), 147–151.
- Cozzi, M. (2018, April). Optimal Capital Taxation with Incomplete Markets and Schumpeterian Growth. Department Discussion Papers 1803, Department of Economics, University of Victoria.
- Croce, M., T. T. Nguyen, and S. Raymond (2021). Persistent government debt and aggregate risk distribution. *Journal of Financial Economics* 140(2), 347–367.
- Croce, M., T. T. Nguyen, S. Raymond, and L. Schmid (2019). Government debt and the returns to innovation. *Journal of Financial Economics* 132(3), 205–225.
- Croce, M. M. M., A. G. Karantounias, S. Raymond, and L. Schmid (2017). A tax plan for endogenous innovation. FRB Atlanta Working Paper 2017-13.

- Erosa, A. and M. Gervais (2002, August). Optimal taxation in life-cycle economies. *Journal of Economic Theory* 105(2), 338–369.
- Gross, T., P. Klein, and M. Makris (2021). Dynamic capital tax competition under the source principle. Forthcoming in *American Economic Journal: Macroeconomics*.
- Howitt, P. and P. Aghion (1998). Capital accumulation and innovation as complementary factors in long-run growth. *Journal of Economic Growth* 3, 111–130.
- Jones, C. I. (1995). R&D-based models of economic growth. *Journal of Political Economy* 103(4), 759–784.
- Jones, C. I. (2019, April). Taxing top incomes in a world of ideas. Working Paper 25725, National Bureau of Economic Research.
- Jones, C. I. and J. C. Williams (1998). Measuring the social return to R&D. *The Quarterly Journal of Economics* 113(4), 1119–1135.
- Jones, L., R. Manuelli, and P. Rossi (1993). Optimal taxation in models of endogenous growth. *Journal of Political Economy* 101, 485–517.
- Jones, L. E., R. E. Manuelli, and P. E. Rossi (1997). On the optimal taxation of capital income. *Journal of Economic Theory* 73(1), 93–117.
- King, R. G., C. I. Plosser, and S. T. Rebelo (2002). Production, growth and business cycles: Technical appendix. *Computational Economics* 20(1), 87–116.
- King, R. G. and S. Rebelo (1990, October). Public Policy and Economic Growth: Developing Neoclassical Implications. *Journal of Political Economy* 98(5), 126–150.
- Lockwood, B. B., C. Nathanson, and E. G. Weyl (2017). Taxation and the allocation of talent. *Journal of Political Economy* 125(5), 1635–1682.
- Long, X. and A. Pelloni (2017). Factor Income Taxation in a Horizontal Innovation Model. *Journal of Public Economics* 154, 137–159.
- Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics* 22, 3–42.
- Lucas, R. E. (1990, April). Supply-side economics: An analytical review. *Oxford Economic Papers* 42(2), 293–316.
- McGrattan, E. R. and E. C. Prescott (2003, May). Average debt and equity returns: Puzzling? *American Economic Review* 93(2), 392–397.
- Ramsey, F. P. (1927, March). A contribution to the theory of taxation. *Economic Journal* 37(145), 47–61.

- Rebelo, S. (1991, June). Long-Run Policy Analysis and Long-Run Growth. *Journal of Political Economy* 99(3), 500–521.
- Reis, C. (2011, June). Entrepreneurial labor and capital taxation. *Macroeconomic Dynamics* 15(03), 326–335.
- Romer, P. M. (1990, October). Endogenous technological change. *Journal of Political Economy* 98(5), 71–102.
- Straub, L. and I. Werning (2020, January). Positive long-run capital taxation: Chamley-Judd revisited. *American Economic Review* 110(1), 86–119.