A Unified Theory of Growth, Cycles and Unemployment - Part I: Technology, Competition and Growth

Pollak, Andreas

University of Saskatchewan

14 December 2022

A Unified Theory of Growth, Cycles and Unemployment Part I: Technology, Competition and Growth*

Andreas Pollak†

Department of Economics, University of Saskatchewan

December 14, 2022

Abstract

Part I of this paper proposes a model of endogenous growth, in which the scale of individual production units is endogenously determined in a novel way. The basic model has desirable growth and static properties, including the following:

(1) The economy exhibits productivity growth at a constant rate that only depends on technology parameters; (2) at the aggregate level, the economy is identical to the neoclassical growth model, thus (3) featuring the full medium-term capital dynamics familiar from this framework; moreover, (4) the model explains why the aggregate production function and many industries exhibit constant returns to scale; (5) there are no unrealistic constraints on the firm-level production technology; in particular, production is not linear in a capital-like input and (6) the notion that R&D investments become less effective with rising technology levels is accounted for; (7) generally, there are no knife-edge conditions or implausible scale effects; (8) no particular assumptions regarding market power or the competitive structure of industries are required; markets can be modelled as perfectly competitive, but the framework is robust to alternative assumptions such as monopolistic competition; (9) being based on the quality-ladder idea, the model can be extended to feature the rich industry-level dynamics that have been studied using Schumpeterian growth models; (10) in its basic version, the model is far simpler while being more general than popular models of endogenous growth.

Keywords: endogenous growth, technology, market structure.

JEL Classification Numbers: O4, O3, E2

*Please check at `http://usask.pollak.org/?research` if an updated version is available. Note that there are two parts to this paper.

†Address: 9 Campus Dr., Saskatoon, SK, S7N 0P9, Canada; telephone: +1 (306) 966-5221; e-mail: *a.pollak@usask.ca*.

1 Introduction

What if we went back to the drawing board? In some areas of macroeconomics, we have been working with the same model frameworks for decades. They have been modified, extended and calibrated to fit a wide range of empirical phenomena, but fundamentally today's theoretical tools are built on the same modelling choices made a long time ago. If we started from scratch today, if we built models based on a few requirements that seem essential given our experience with existing theories, what would we get? This paper asks this question for the areas of endogenous growth and business cycles. As we will see, the answer is that we arrive at models that are, in some important ways, quite different from what we are using today, and that there is a lot we can learn from such a model building exercise.

After rapid progress and a surge in publications in the area of endogenous growth in the 1990s and early 2000s that led to the development of many of the currently popular theoretical frameworks,[1] the focus of the literature has shifted towards applications of existing theories to a wide variety of issues, many of them related to industry-level dynamics, while building on the tenets that had emerged from the earlier theoretical work.

There appears to be a broad consensus that growth is made possible by the directed and profit-driven efforts of firms and entrepreneurs, and that the factors determining the cost of innovation and the returns that can be realized by developing new technologies and products matter. However, the workhorse models used since the 1990s still have several well-known challenges. These include knife-edge conditions, meaning that functional forms and model parameters must be chosen very carefully to avoid unrealistic unstable behaviour. Many of the earlier models of endogenous growth involved scale effects, where economic outcomes depend on the size of the economic system.[2] Theories often rely on very specific

[1]See Jones (2019) on Romer's contribution that sparked renewed interest in questions surrounding the sources of growth.

[2]Dinopoulos and Thompson (1999) provide an early overview of the issue of scale effects as they pertain to Schumpeterian models.

assumptions regarding the competitive structure in individual markets or even strategic interactions between firms, aspects that macroeconomic models in other areas are commonly able to abstract from. While many of these issues have been solved in principle, there is no widely used framework that gets close to the neoclassical growth model in terms of its simplicity and consistency with important features of the aggregate economy. Despite their great success in explaining a wide range of phenomena,[3] endogenous growth models appear lack a degree of generality and robustness that would allow us to derive results that could be expected to hold in a variety of plausible settings and in changing environments.

Popular business cycle theories are equally able to match many aspects of the data rather well, and just like models of endogenous growth, they have grown increasingly complex. However, insofar as they build on real business cycle (RBC) theory[4] or are constructed based on similar principles and ideas, they are vulnerable to many of criticisms that have been levelled agains the use of RBC models.[5] To the extent that productivity shocks are assumed to be the source of fluctuations, one open question is how we can rationalize and explain variation of the required magnitude, especially negative changes.[6] Moreover, such models often lack powerful amplification and propagation mechanisms or an internal structure that can provide a good explanation for the clear asymmetry we observe in business cycle patterns. The magnitude and persistence of fluctuations is usually largely dictated by the characteristics of the external shocks. This makes RBC models and similar frameworks rather "passive" in a sense. We rattle the economy and observe what parts shake and how, and we ask how we can dampen some of the movements, but little of the action is endogenous to the model. As such, many of the popular business cycle models do not explain the source and nature of cycles or recessions, they merely describe how the economy responds to them.

[3]An example is the application of quality-ladder models due to Grossman and Helpman (1991) and Aghion and Howitt (1992) to issues of growth and firm dynamics; see Klette and Kortum (2004), Acemoglu and Cao (2015), Acemoglu, Akcigit, et al. (2018), and Akcigit and Kerr (2018).

[4]Kydland and Prescott (1982) and Prescott (1986).

[5]The New Keynesian framework is the most widely used approach directly building on the ideas of the RBC model. See Woodford (2003) for an excellent introduction and Galí (2018) for a recent summary.

[6]See Summers (1986) for an early criticism of aspects of the RBC model. Mankiw (1989) discusses shortcomings of the RBC framework compared to the emerging New Keynesian theories.

This paper argues that by introducing some relatively simple changes to how we model the firm sector – effectively adding an additional layer to the neoclassical growth model – we arrive at a description of the aggregate economy that can be the basis for both endogenous growth and business cycle models that are more general and robust while at the same time being simpler than many of the popular frameworks. The paper is split into two parts. This first part focuses on issues related to endogenous growth.

In using virtually the same model to explain both growth and business cycle phneomena, our approach follows the agenda at the core of the real business cycle literature that highlights the importance of being able to explain both long-term and short-term outcomes using the same model framework. Prescott (1986) comments that "[t]he use of the expression *business cycle* is unfortunate [...]. [I]t leads people to think in terms of a time series' business cycle component which is to be explained independently of a growth component; our research has, instead, one unifying theory of both of these."[7] This paper aims to make further progress towards a unified theory of growth and cycles by being explicit about the endogenous factors determining technological progress and highlighting the role of involuntary unemployment as an important aspect of cyclical macroeconomic outcomes.

1.1 Model Idea

One of the most stunning macroeconomic patterns is the stability of the rate of productivity growth in economies near the technological frontier. For the past 150 years at least – as long as reliable records on output go back – the growth rate has remained close to 2% per year.[8] This observation is particularly surprising considering how much the world has changed during this long time span.

Demographics are very different now from what they were in the past, with population growth dropping dramatically during the 20th century from its previously high levels. The

[7] Ibid., p. 12.

[8] This point has been made a in particular by proponents of semi-endogenous growth models, see for example Jones (1995b) and Jones (2002).

openness of economies to international trade and investment has moved from largely unrestricted in the 19th century to protectionist in the first half of the 20th century and back to rather globalized today. The part of the world that participates in the free exchange of goods, assets and ideas is much larger today than in the decades following World War II. The sectoral composition has shifted with agriculture giving way to manufacturing and services, and with it many characteristics of the goods on which expenditure is concentrated. At the same time, the concentration and exposure to international competition has changed in many industries with likely effects for market power. In addition to all this, there is evidence that developing the technological improvements that fuel economic growth has become more difficult and costly over time.[9]

If we accept the notion that productivity-relevant research and development (R&D) is conducted by firms seeking profits or at least market returns on their R&D investments, and realize that these returns are linked to the firms' sales, we have to ask: How is it possible that these returns warrant R&D investments that lead to stable growth, even though the size of economies and the relevant world market as well as the cost of these investments have changed considerably?

The answer is that we must differentiate between incentives at the firm level and aggregate phenomena. As long as there is a mechanism that guarantees that individual firms can always earn a market return on their R&D investments, because either the scale of the sales or the mark-ups associated with particular investments are able to adjust appropriately as the resources required to develop and implement a given technology improvement change, firms will be able to generate technology improvements at a constant rate.

We will refer to the production tied to an individual stream of R&D investment as a project. As output of individual projects is linked to characteristics of the knowledge production function for a given rate of growth and thus determined independently of the

[9]Kortum (1993) discusses apparently contradicting evidence on patent creation and R&D expenditure data. Segerstrom (1998) and Kortum (1997) develop theoretical models to explain these observations and conclude that it must be getting harder over time to innovate. Klette and Kortum (2004) model the relationship between firms' R&D expenditure, firm size and growth.

overall market size in an industry, it must be possible to freely create and terminate individual projects. For example, in an industry where the cost of delivering technological improvements grows faster than the size of the industry itself, there must be ongoing concentration, with the number of active projects shrinking over time while their individual scale is growing faster than the industry. The production factors freed up as projects disappear can be repurposed for the production of continuing projects, either in the same or a different industry.[10]

Another implication of allowing for R&D to improve products or productivities is increasing returns to scale at the level of projects. It seems plausible to assume that *statically*, projects in most industries should be able to operate under at least constant returns to scale. Having the option to invest in R&D then adds an additional production factor to the technology when *dynamic* adjustments are considered, resulting in increasing returns to scale. We will refer to this dynamic production technology with increasing returns to scale as *atemporal*, because it includes production techniques that are used at different points in time and that involve different levels of technology and knowledge.

With increasing returns to scale at the project level, it appears that we are dealing with a natural monopoly situation, where competition among firms is not only unstable, but also undesirable from the perspective of productive efficiency. It is important to remember, though, that these scale effects result from the possibility to invest in new ideas, and that the resulting production technology is an atemporal one. The question is therefore not simply: Why doesn't a single firm take advantage of scale effects to take over the market? It is rather: Why doesn't a single firm use future technology and knowledge to improve its productivity and take over the market? The answer to this is question obvious. While it is certainly possible to accelerate R&D, the accumulation of knowledge and the rate of productivity

[10]The scale-independent growth models of Young (1998) and Dinopoulos and Thompson (1998) are based on the same idea that firm size should generally be independent of market size and that additional entry and the resulting duplication of R&D efforts is the mechanism that absorbs scale effects. The relationship of our approach to these papers will be discussed in section 2.7 below. Klette and Kortum (2004) similarly present a model in which the scale and scope of firms is allowed to vary in principle. However, their focus, and similarly the focus of the work that builds on their model such as Akcigit and Kerr (2018), is more on firm-level dynamics than aggregate and growth properties.

growth, doing so is costly. We will account for this cost of moving towards better production techniques at a faster pace, which will help address the static market structure problems that are often associated with increasing returns.

With this, all ingredients of our model are in place. From the assumption that R&D drives productivity advances and the requirement that growth should have the potential to be independent of demographics and market size, we get the three fundamental model assumptions of free entry and exit, increasing returns at the project level and costs associated with the acceleration of productivity improvements. We will see that the resulting model makes predictions that go far beyond the immediate and obvious implications of these model assumptions.

1.2 Plan of this Paper

The following section 2 starts by developing a model of endogenous growth. We will begin with a project-level production function that is characterized by increasing returns to scale in the long term but otherwise is entirely standard. From there on, we will build the model up to a representation of the aggregate economy, which will be identical to the production side of the canonical neoclassical growth model with exogenous technological progress. The aggregate production function will have constant returns to scale, stable factor shares and TFP growth, and exhibit the short to medium term adjustment dynamics familiar from the Solow-Swan[11] or Ramsey-Cass-Coopmans[12] models of exogenous growth. We will highlight some features of our model regarding the role of market power and the appropriability of R&D and position our framework in relation to the existing growth literature.

In section 3, we will look at how our model can provide new insights into issues related to growth theory or in macroeconomics in general, and consider some possible extensions. Finally, section 4 summarizes our findings.

[11]Solow (1956), Solow (1957), and Swan (1956).
[12]Ramsey (1928), Cass (1965), and Koopmans (1963).

2 Economic Growth

The starting point for modelling aggregate economic growth will be to track production and productivity at the level of individual projects. We will think of a project as a possibly differentiated product or product group that is produced using dedicated production factors, capital and labour, where the appropriate definition of capital is broad enough to encompass intellectual property, including that resulting from investment in R&D. In some industries, each project may be associated with an individual firm[13], while in others firms may host many projects[14]. Similarly, it is possible for multiple firms to jointly produce the output associated with a single project[15].[16]

Suppose the production technology for an individual project is given by the Cobb-Douglas production function $f(k, n) = Dk^{\alpha}n^{\beta}$ for positive constants D, α and β. The source of growth at the aggregate level will be increasing returns to scale in the two production factors capital k and labour n at the level of individual projects, i.e. $\alpha + \beta > 1$. Increasing returns, while not the most common modelling choice in much of the macroeconomic literature in general, are an essential feature of many endogenous growth models. Intuitively, if we are willing to assume that statically, the production technology exhibits constant returns to scale in (narrowly defined) capital and labour, but at the same time productivity-enhancing investment in total factor productivity (TFP) is possible, the technology has, in fact, increasing returns to scale in all three factors TFP, capital and labour combined. The factor TFP, being the result of prior investment in technology, can then be interpreted as a different type of capital, which we will include in our broad definition of capital k. Appendix A motivates the use of this broad definition of capital.

Arguably, the assumption of increasing returns is highly plausible for R&D intensive

[13]Family-owned restaurants may be an example of a one-for-one match between projects and firms.

[14]Most firms that offer multiple product lines would fall into this category.

[15]Examples include franchise arrangements, the licensing of product designs to multiple manufacturers, generally vertically disintegrated production processes, and, as a more recent example, ride service platforms.

[16]The distinction between projects and firms is not relevant for any of the mechanisms related to growth or business cycles we discuss below. It is, however, important for the interpretation of some of the model mechanics.

industries, where large R&D expenditures correspond to significant fixed costs when production is looked at statically. Increasing returns at the firm level are often associated various modelling challenges, from the inability to pay factors according to their marginal product to issues around market structure. We will see below that such problems do not arise in our setting, and that thanks to free entry and costs associated with moving towards better production techniques, firm behaviour and outcomes will be similar to those under the simple benchmark of perfect competition with constant returns to scale.[17]

Much of the literature on endogenous growth assumes constant returns to capital, $\alpha = 1$. Popular frameworks employ an aggregate production function that is linear in some measure of capital[18], which can be expanded at a cost that is proportional to the current productivity level. An exception is the AK model[19], where the distinction is made between individual firms operating at constant returns to scale and an aggregate technology that has constant returns to capital thanks to a well-calibrated externality. The reason why constant returns to capital are considered necessary for growth to happen is to avoid the trap of diminishing marginal products, which would lead the economy to converge to a stable steady-state level of capital that only depends on the scale of the fixed factor.

Of course, both the case of constant returns to scale $\alpha + \beta = 1$ and the case of constant returns to capital $\alpha = 1$ impose very specific constraints on the aggregate production technology that are exceedingly unlikely to arise by chance.[20] If they are relevant in practice, it would be desirable have them emerge endogenously in a robust fashion from some plausible mechanism. The middle ground of $1 > \alpha > 1 - \beta$ is occupied by the literature on semi-endogenous growth.[21] While there are increasing returns to scale at the aggregate level, for

[17] Basu and Fernald (1997) point to the inconsistency between the scale elasticities measured at the firm or industry level compared to economy-wide aggregates, and conclude that the firm sector is not described well by a single representative firm. Ahmad, Fernald, and Khan (2019) find that at the aggregate level, the US economy is characterized by constant returns to scale.

[18] e.g. the varieties model or the Schumpetrian model

[19] See P. M. Romer (1986).

[20] The former case of constant returns to scale at the aggregate level is far easier one to justify, for example as the production technology in an economy that is so large that all scale effects have been exhausted.

[21] See Jones (1995a) and Jones (2005).

any level of the fixed factor labour, there is an optimal level of capital. Persistent growth is only possible to the extent that this labour input grows over time.

Our framework is geared towards the same parameter range of $1 > \alpha > 1 - \beta$ as semi-endogenous growth models. However, in contrast to that framework, we associate this technology with individual projects rather than the aggregate production function, and we break the link between aggregate labour supply and growth by allowing firms to grow independently at an appropriate rate.[22]

2.1 Production and Growth Acceleration Costs

Let the technology at the level of an individual project be given by

$$f(k, n) = Dk^{\alpha}n^{\beta} \tag{1}$$

for factor inputs capital k and labour n with $0 < \alpha < 1$, $0 < \beta < 1$ and $\alpha + \beta \geq 1$.

Broad capital k includes, besides physical capital, the value of productivity-enhancing technology to the extent that it is appropriable and that it contributes to output in a way that is consistent with the production function (1).[23] Capital depreciates at the rate δ.

The production function f is *atemporal* in the sense that it comprises of production techniques that require different levels of knowledge. It encompasses the best ways of producing output at different stages of development of an economy. The higher capital and labour inputs associated with higher productivity are not merely larger quantities, they will typically include higher quality of physical and human capital as well as intangible assets that represent a higher level of technology. More productive techniques are therefore generally

[22]This can work as long as firms can grow at a rate that is different from the aggregate growth rate. Once concentration in each industry is so high that no further exit is possible, the constraints highlighted in the semi-endogenous growth literature apply.

[23]As seen in appendix A, these conditions are met if it is possible to invest in total factor productivity improvements under isoelastic investment costs and if the resulting technological knowledge or the capital that embodies it can be traded between firms. Note that to the extent that technological knowledge is not fully appropriable, this would be reflected in a higher depreciation rate and a lower cost of acquisition of this particular asset.

not immediately available to firms. A car manufacturer in 1950 could not simply have chosen to use a 1980 production technique to produce vehicles of unrivalled quality at extremely low cost and take over the market this way. We will assume that while techniques that have previously been employed by some producer are readily available for implementation by anyone, the use of new techniques requires the creation of new knowledge and the modification of a previously known production process, all of which is associated with a cost. This cost plausibly increases with the rate at which the production technique is changed, which is why we will refer to it as a growth acceleration cost (GAC).

We will assume that the GAC is multiplicatively separable into a component related to scale of a firm's R&D activity and a component ψ that only depends on the relative change of the production technique. Let $(\bar{k}, \bar{n})$ be the factor input associated with the "best" previously tried production technique. A firm planning to move to the factor input combination (k, n) will incur a relative cost $\psi = \psi(\frac{k}{\bar{k}} - 1, \frac{n}{\bar{n}} - 1)$. If speeding up R&D only makes the acquisition of new knowledge capital more expensive, this cost would scale with capital for an absolute GAC of ψk. If, on the other hand, the creation of new knowledge and ideas requires inputs similar to production, or if the development or implementation of new production processes disrupts production or competes with it for resources, the GAC is tied to output instead, $\psi f(k, n)$.

Appendix B shows that these two specifications of the cost function are practically almost indistinguishable. We will model GACs as being tied to output, which gives them the flavour of an opportunity cost. The instantaneous output of a project net of these costs is then $y = \left(1 - \psi(\gamma^{[k]}, \gamma^{[n]})\right) f(k, n)$ with $\gamma^{[k]} = \frac{k}{\bar{k}} - 1$, $\gamma^{[n]} = \frac{n}{\bar{n}} - 1$.[24]

This GAC ψ creates a discrepancy between the long-term production technology f and the relevant production possibilities available to the firm in the short run, which are given

[24]Formally, the GAC is reminiscent of the concept of adjustment costs, popularized in the 1960s by Lucas (1967) among others, which is a standard feature of many macroeconomic and finance models, including importantly the q-theory of investment, due to Kaldor (1966) and Tobin and Brainard (1976). Note however, that our interpretation is quite different, and that we will be mostly interested in general equilibrium implications of these costs that are not tied to short-term fluctuations in capital intensities.

by $(1-\psi)f$ and depend on the set of previously tried factor input combinations. We will assume that the function ψ is strictly convex, twice differentiable and has a global minimum at $\psi(0,0) = 0$. We also allow for the special case that ψ only depends on one of its arguments. There is no cost associated with using a known technique, $\psi(0,0) = 0$, and marginal adjustments to a known production process are free, but costs generally increase superlinearly with the relative adjustment in factor inputs and become prohibitive for sufficiently large increases in at least one factor.

As discussed above, there is free entry and exit. Firms can create additional projects or shut down existing ones, new firms can enter with their own projects and then exit at any time. If a new project is created, it needs to be endowed with the desired amount of production factors. When a project is terminated, the factors previously tied up in its production are released. Free entry and exit ensure that the profits associated with projects are zero. This implies that firms, which are collections of individual projects, earn zero profits as well.

The firm's problem is static, albeit history dependent due the importance of previously used production techniques for GACs. For a known production technique $(\bar{k}, \bar{n})$, a firm maximizes the profit associated with a particular project

$$\pi = \left(1 - \psi(\frac{k}{\bar{k}} - 1, \frac{n}{\bar{n}} - 1)\right) f(k,n) - wn - (r+\delta)k \tag{2}$$

taking the current wage w and interest rate r as given.

Letting f_k and f_n be the marginal products of capital and labour, respectively, as implied by the atemporal production function f and similarly ψ_k and ψ_n the marginal GACs, the first-order conditions for the optimal choice of capital and labour immediately imply the following:

$$r + \delta = (1-\psi) f_k - \psi_k \frac{f(k,n)}{\bar{k}} \tag{3}$$

$$w = (1-\psi) f_n - \psi_n \frac{f(k,n)}{\bar{n}} \tag{4}$$

If firms adjust the factor inputs associated with projects, they pay factor prices *below* the respective long-term marginal products $(1-\psi)f_k$ and $(1-\psi)f_n$. This makes it possible for firms to earn nonnegative profits even if increasing returns to scale in f imply $(1-\psi)f_k k + (1-\psi)f_n n > (1-\psi)f(k,n)$.

Proposition 1 (Zero Profit Optimum) *Consider an allocation that satisfies both the first-order necessary conditions for the optimal choice of factor inputs* (3) *and* (4) *and the zero-profit condition* $\pi = 0$ *for positive factor inputs. This allocation has the following properties.*

1. Optimality: *The allocation is a profit maximum.*

2. Uniqueness: *For any given factor input ratio, there is at most one such allocation.*

3. Scale Elasticity: *At this allocation, the instantaneous production function* $(1-\psi)f$ *locally exhibits constant returns to scale.*

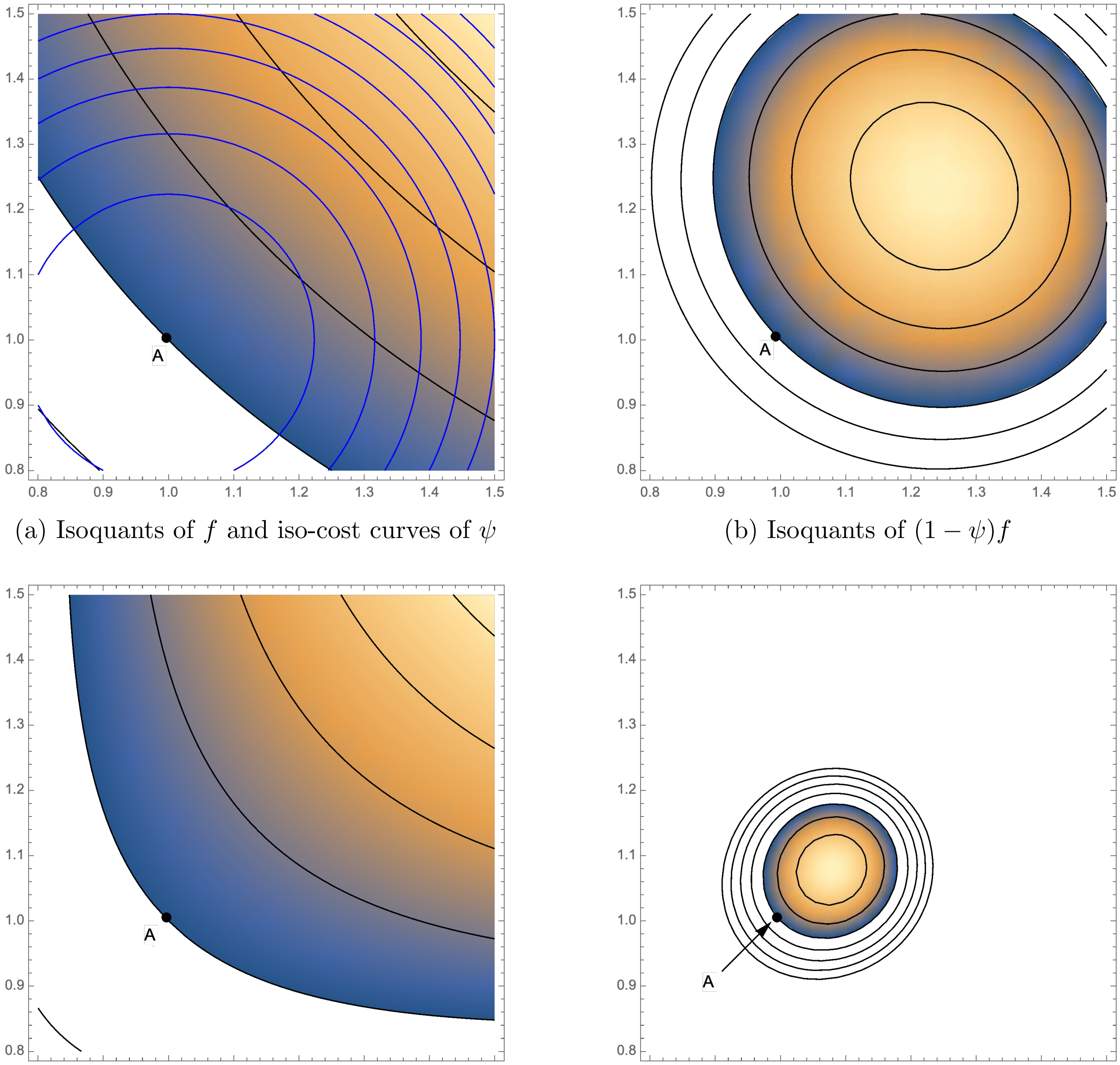
These results generalize to the case where ψ *only depends on one factor input,* $\psi(\gamma^{[k]}, \gamma^{[n]}) = \psi^{[i]}(\gamma^{[i]})$ *for* $i \in \{k, n\}$ *and* $\psi^{[i]}$ *strictly convex with* $\psi^{[i]}(0) = \psi^{[i]\prime}(0) = 0$*.*

Proof. In appendix C. ■

We can use (3) and (4) to eliminate the factor prices from the free-entry condition $\pi = 0$. The resulting zero-profit condition links the rate of technological adjustment to the scale elasticity of f.

$$\frac{\psi_k \frac{k}{\bar{k}} + \psi_n \frac{n}{\bar{n}}}{1-\psi} = \frac{f_k k}{f(k,n)} + \frac{f_n n}{f(k,n)} - 1 = \alpha + \beta - 1 \tag{5}$$

Figure 1: Production Technique and Growth Acceleration Costs



(a) Isoquants of f and iso-cost curves of ψ

(b) Isoquants of $(1-\psi)f$

(c) Iso-profit lines for f

(d) Iso-profit lines for $(1-\psi)f$

Note: Using production function $f(k,n) = k^{\frac{2}{3}} n^{\frac{2}{3}}$ and costs $\psi(\gamma^{[k]}, \gamma^{[n]}) = \left(\gamma^{[k]}\right)^2 + \left(\gamma^{[n]}\right)^2$. The profit is based on factor prices that are chosen such that point A is efficient and zero-profit. See text for further details.

With constant returns to scale at the project level in the long run, $\alpha + \beta = 1$, there is neither room nor an incentive to devote resources to advancing the production technique, and at least as long as factor prices do not change, alternative factor input combinations remain unexplored. With increasing returns to scale, firms can and will improve their productivity over time.

Even if the atemporal production technology f has increasing returns to scale, the cost ψ guarantees that individual firms cannot get arbitrarily productive in any given period. The properties of the cost function ψ ensure that for a large enough increase in factor inputs, $\psi = 1$ and therefore $y = 0$. There is a maximum amount of output that firms can reach within a period by adjusting the production process. Considering factor costs, the optimal output is below this maximum output. Yet, because ψ and its derivatives remain arbitrarily close to zero in a neighbourhood of $(\bar{k}, \bar{n})$, firms always have an incentive to increase their factor inputs somewhat as long as scale effects exist. The combination of local scale effects for any previously used production technique with an accelerating cost of implementing changes to the production process that is bound to dominate any productivity improvements for larger increases in factor inputs guarantees positive but finite productivity growth.

Figure 1 illustrates the role of the GACs ψ for the choice of the production technique within a period. In all panels factor inputs are shown on the two axes. Point A at $(1, 1)$ represents the best previously known technique. Some isoquants associated with the production technology f are shown as black lines in panel 1a. In addition to that, the area of improvement compared to point A is highlighted with a coloured height map. Due to increasing returns to scale in f, the firm could, for given factor prices and in the absence of GACs, increase its profit arbitrarily by increasing both factor inputs. This is shown in panel 1c, which displays the iso-profit lines associated with f for the factor prices that make point A efficient and zero-profit.

The blue circles in 1a are iso-cost curves representing the GAC ψ. This example uses a quadratic cost function based on the Euclidean norm. The implication of this cost is that choosing factor input combinations further away from A reduces output, with the effect becoming prohibitive for large deviations. Panel 1b shows the isoquants of the instantaneously available technology $(1-\psi)f$ in the current period. The set of factor input combinations that result in improved output in the current period is now bounded and there is a maximum level of output that can be reached. Taking into account factor costs, the efficient level of output is somewhat below that maximum level, as can be seen in panel 1d, which displays iso-profit curves taking into account GACs.

Equations (3) and (4) make it clear that the relationship between factor inputs and factor prices depends on the properties of the function ψ. How its properties relate to the static properties of an economy as well as its growth path is discussed in more detail below.

2.2 The Basic Model

We start by considering a special case that is both intuitively appealing and very simple. Suppose all R&D efforts result in the creation of knowledge and ideas that form part of a firm's value or can at least be proxied for by its assets, and that a firm's technological level is independent of the labour intensity ultimately chosen in production.[25] In this case, the GAC only depends the growth rate of capital $\gamma^{[k]}$ at the firm level, not on employment adjustments, so that $\psi(\gamma^{[k]}, \gamma^{[n]}) = \psi(\gamma^{[k]})$. R&D expenditure has two components. R&D *investment* leads to the creation of – possibly intangible – assets, the value of which reflects the resources required to acquire a similar asset again later. The *cost* ψ is the expenditure needed to develop these assets before their characteristics become common knowledge. The more advanced these assets are compared to the prior technology level indexed by $\bar{k}$, i.e. the higher the rate of technology improvement, the higher is this cost.

Figure 2 illustrates this case of univariate GACs. Even though costs only grow with one production factor (in the direction of the abscissa in figure 2a) and firms could thus, in

[25] This notion is consistent with how R&D is modelled in the majority of the endogenous growth literature, where it is usually assumed that any R&D expenditure is tied directly to a TFP or labour productivity outcome.

principle, produce arbitrarily high output by increasing the other input (vertical movement in 2b), the optimal adjustment of the production technique within a period is still always finite. With the increase of one factor constrained by GACs, the optimal increase of the other factor is, for given factor prices, limited by its diminishing marginal product (see 2d).

With this assumption, equation (5) simplifies to

$$\frac{\psi'(\gamma^{[k]})(1+\gamma^{[k]})}{1-\psi(\gamma^{[k]})} = \alpha + \beta - 1. \tag{6}$$

The properties of the cost function ψ guarantee that this equation has a unique solution for a rate of capital growth that is both nonnegative and admissible, $0 \leq \gamma^{[k]} \leq \psi^{-1}(1)$, which depends on the specification of the cost function and the scale elasticity of f, $\alpha + \beta$.

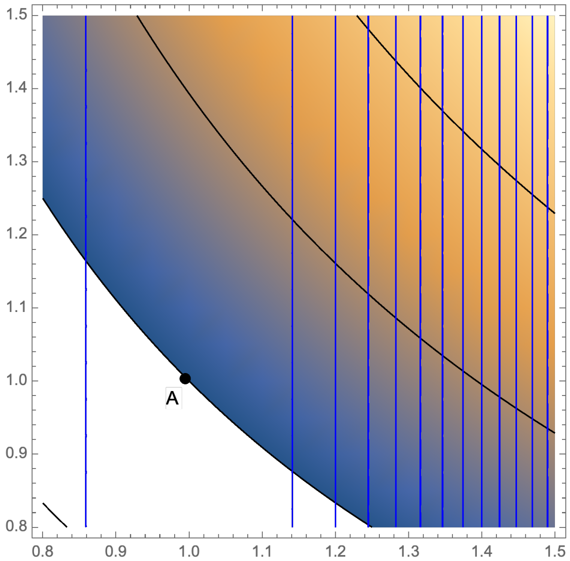
With $\psi_n = 0$, equation (4) implies a labour share of β, $wn = (1-\psi)f_n n = \beta y$. Zero profits then require a capital share of $1 - \beta$.

Let the capital input per project at time $t = 0$ be k_0. Because of the fixed growth rate of the project-level capital stock $\gamma^{[k]}$, capital in period t is then $k_t = k_0(1+\gamma^{[k]})^t$. For an aggregate capital stock K_t and an economy-wide labour supply N_t in period t, there are $\frac{K_t}{k_t}$ projects each operating the production technology f with capital k_t and labour $n_t = \frac{k_t}{K_t} N_t$. Aggregate output in the economy is thus given by

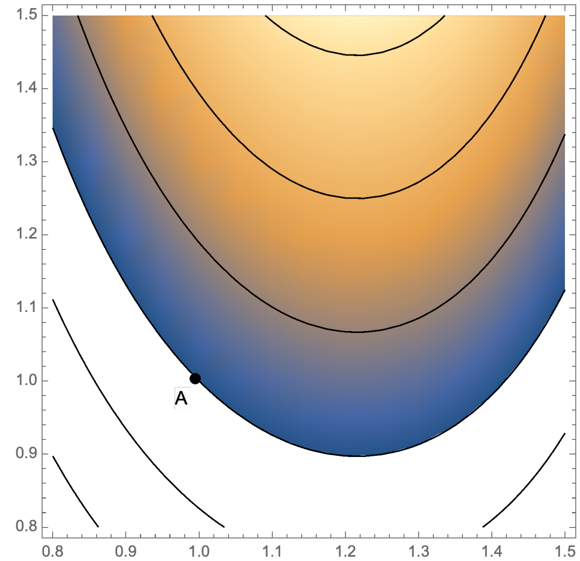
$$\begin{aligned} Y_t =& \frac{K_t}{k_t} y_t = \frac{K_t}{k_t}(1-\psi) D k_t^{\alpha} n_t^{\beta} \\ \doteq& (1-\psi) D k_0 \left(1 + (\alpha+\beta-1)\gamma^{[k]}\right)^t K_t^{1-\beta} N_t^{\beta}. \end{aligned} \tag{7}$$

Here and in what follows, we use the symbol $\doteq$ to denote an approximate equality that is exact in the continuous-time limit, where the period length becomes zero. If we define the total factor productivity as $A_t = (1-\psi) D k_0 \left(1 + (\alpha+\beta-1)\gamma^{[k]}\right)^t$, the economy can be described at the aggregate level by the standard, constant-returns-to-scale production technology $Y_t = A_t K_t^{1-\beta} N_t^{\beta}$, where total factor productivity A_t grows at a constant rate, as long as the project-level atemporal technology exhibits increasing returns to scale, $\alpha + \beta > 1$.

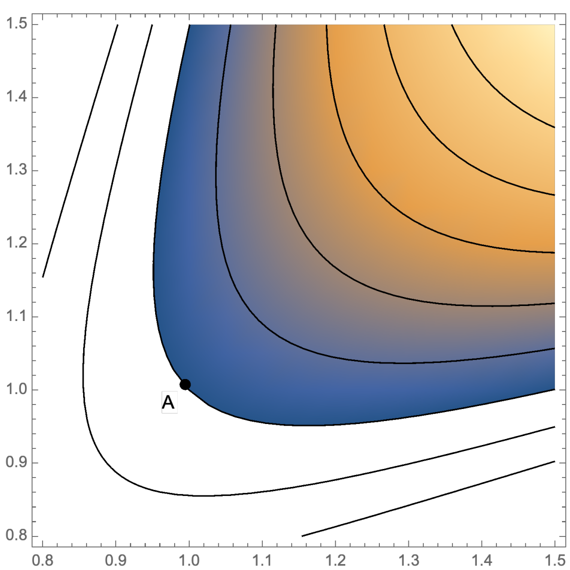
Figure 2: Production Technique and Growth Acceleration Costs for a Single Factor Only



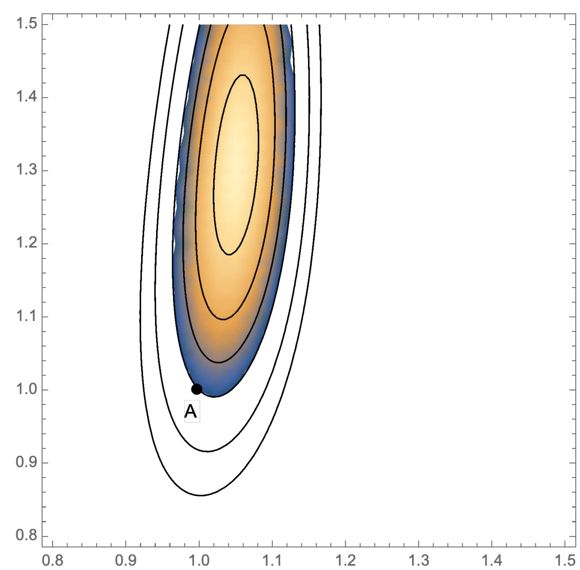
(a) Isoquants of f and iso-cost curves of ψ



(b) Isoquants of $(1-\psi)f$



(c) Iso-profit lines for f



(d) Iso-profit lines for $(1-\psi)f$

Note: Using production function $f(k,n) = k^{0.55}n^{0.55}$ and costs $\psi(\gamma^{[k]}, \gamma^{[n]}) = \left(\gamma^{[n]}\right)^2$. The profit is based on factor prices that are chosen such that point A is efficient and zero-profit. See text for further details.

Moreover, capital and labour income are consistent with their marginal products in the aggregate production function, $r_t + \delta = (1-\beta)A_t K_t^{-\beta} N_t^{\beta}$ and $w_t = \beta A_t K_t^{1-\beta} N_t^{\beta-1}$, and the aggregate capital and labour shares are thus constant at the expected levels of $1-\beta$ and β.

The aggregate economy behaves in every respect exactly like the neoclassical growth model for a constant rate of productivity growth. This means that under the usual assumptions regarding households' preferences or behaviour, there is a stable balanced growth path involving a well-defined ratio of capital K to effective labour $A^{\frac{1}{\beta}} N$, and the familiar adjustment dynamics kick in whenever this ratio deviates from its steady-state value.

There is no particular reason why at any time capital growth at the project level $\gamma^{[k]}$ should be equal to the growth rate of the aggregate capital stock $\gamma^{[K]}$. This means that there is, in general, ongoing entry of new projects or the exit of existing ones with a project termination rate of $\tau \doteq \gamma^{[k]} - \gamma^{[K]}$. In a steady state involving a constant interest rate, aggregate capital must grow at the rate $\gamma^{[K]} = \frac{1}{\beta}\gamma^{[A]} + \gamma^{[N]} = g + \gamma^{[N]}$, where $\gamma^{[A]} = (\alpha + \beta - 1)\gamma^{[k]}$ is TFP growth, $g = \frac{\alpha+\beta-1}{\beta}\gamma^{[k]}$ is the growth rate of labour productivity and $\gamma^{[N]}$ is the growth rate of aggregate employment. This results in a steady-state project termination rate of $\tau \doteq \frac{1-\alpha}{\beta(\alpha+\beta-1)}\gamma^{[A]} - \gamma^{[N]} = \frac{1-\alpha}{\alpha+\beta-1} g - \gamma^{[N]}$, which is positive for plausible model parameterizations.[26]

We will consider out-of-steady-state project termination dynamics in part II of this paper, where they will play a role in investment booms and the slow recovery from recessions.[27]

2.3 More General Cost Functions

In what follows, we generalize the GAC function ψ to the bivariate case. In order to provide the intuition for the relevant model mechanics, we first consider a scenario involving a functional form of ψ that is simple enough to derive an approximate analytical solution

[26] For $\alpha \leq \frac{2}{3}$, $\alpha + \beta \leq \frac{4}{3}$ and $g = 2\%$, we have $\tau \geq 2\% - \gamma^{[N]}$.

[27] In the otherwise completely symmetrical case where GACs only apply to changes in project-level labour input, $\psi(\gamma^{[k]}, \gamma^{[n]}) = \psi(\gamma^{[n]})$, the termination rate $\tau \doteq \gamma^{[n]} - \gamma^{[N]}$ is independent of capital adjustment dynamics and only depends on aggregate labour supply growth.

before characterizing outcomes for a more general class of convex cost functions.

Let $\psi(\gamma^{[k]}, \gamma^{[n]}) = a\left(\gamma^{[k]}\right)^2 + b\left(\gamma^{[n]}\right)^2$ for constants $a, b \geq 0$, $a + b > 0$. We will be interested in situations where both factors grow at nonnegative but low rates and the period length is short. In this case, we expect $\frac{k}{\bar{k}}$, $\frac{n}{\bar{n}}$ and $1 - \psi$ to be close to one, so that we can approximate the free-entry condition (5) as

$$a\gamma^{[k]} + b\gamma^{[n]} = \frac{\alpha + \beta - 1}{2}. \tag{8}$$

This equation characterizes the set of factor input combinations for which profits in the firm sector are zero. It is shown as the downward sloping line in figure 3.

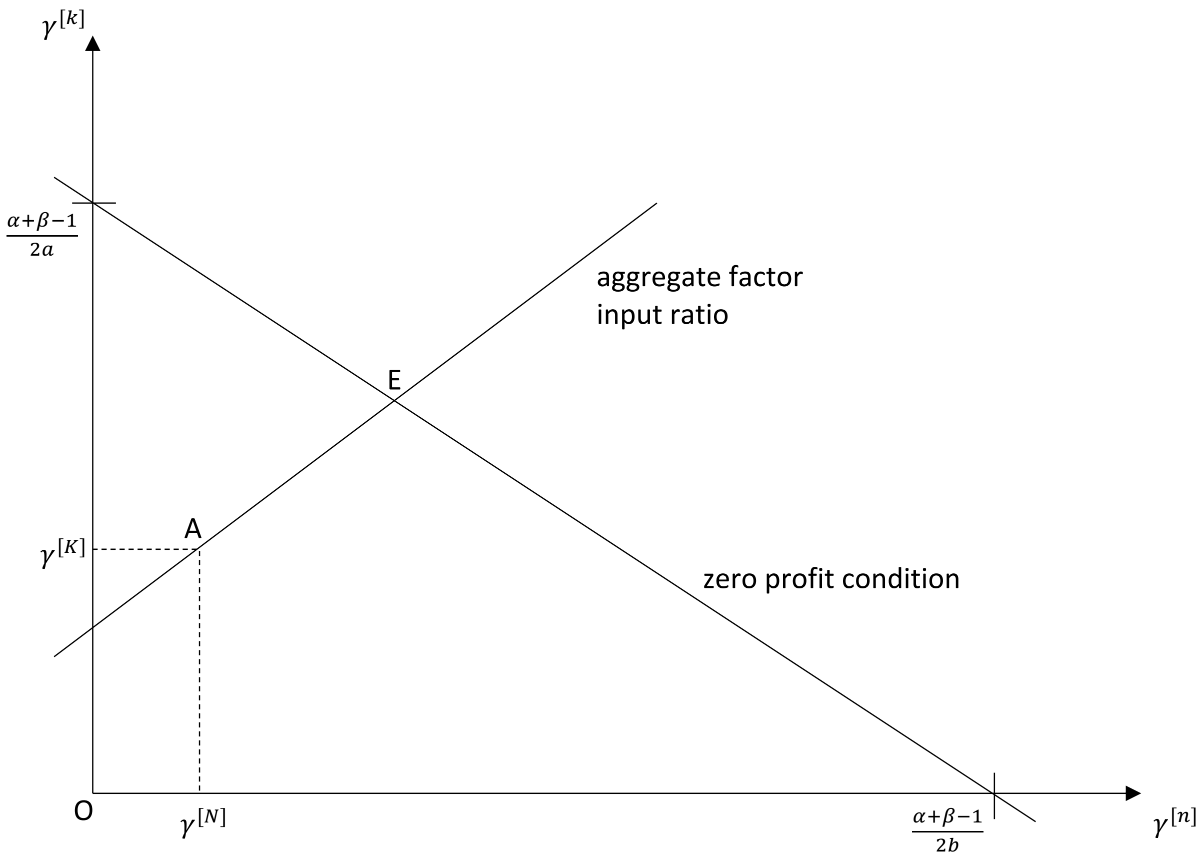
For a point on this line to be feasible, it must be consistent with available factor inputs.

The origin in figure 3 represents the previous factor input combination. Suppose that both aggregate capital and aggregate labour have grown at positive rates, $\gamma^{[K]}$ and $\gamma^{[N]}$, respectively. If each individual project increased its factor inputs at exactly these rates, factor markets would clear. This would be the case at point A in the figure. While this allocation is feasible, it is not consistent with the free entry condition. Marginal GACs would be so low that firms would lose money if they paid both factors their marginal product. At the same time, scale effects would provide an incentive to grow.

Reaching a feasible allocation that is consistent with the free entry condition requires project exit. If projects are terminated at a rate of $\tau = 1\%$, for example, both project-level factor input growth rates can exceed their aggregate counterparts by one percentage point. The unit-slope line labelled "aggregate factor input ratio" in figure 3 shows all feasible allocations. Points to the top right of A require project terminations, $\tau > 0$, those to the bottom left are associated with entry, $\tau < 0$. The equilibrium E is the allocation that guarantees both zero profits and market clearing in both factor markets.

As available factor inputs and thus $\gamma^{[K]}$ and $\gamma^{[N]}$ are predetermined at any point in time, the equilibrium E, and with it the change in the number of active projects, is determined

Figure 3: Equilibrium Determination



statically as the solution to (8) and the aggregate constraint $\gamma^{[k]} - \gamma^{[n]} \doteq \gamma^{[K]} - \gamma^{[N]}$. The factor input growth rates at the project level determine the empirical growth rate of total factor productivity. Thanks to the linearity of the free-entry condition (8), aggregate output is still consistent with a stable Cobb-Douglas production function $Y_t = A_t K_t^{\bar{\alpha}} N_t^{1-\bar{\alpha}}$ with constant TFP growth *independent* of the actual growth rates of factor inputs, although factor income shares may vary. The capital coefficient $\bar{\alpha} = \frac{a(1-\beta)+b\alpha}{a+b}$ is the weighted average of the values $1-\beta$ and α that arise if only one factor is subject GACs. The rate of TFP growth works out to $\frac{1}{2}\frac{(\alpha+\beta-1)^2}{a+b}$.[28,29]

[28]This result is most easily derived by rewriting aggregate production $Y = y\frac{N}{n}$ in growth rates and then using equation (8) and the factor market clearing condition $\gamma^{[k]} - \gamma^{[n]} \doteq \gamma^{[K]} - \gamma^{[N]}$ to eliminate the firm-level growth rates $\gamma^{[k]}$ and $\gamma^{[n]}$.

[29]All of these results are, of course, approximate, with the accuracy of the approximation being better

The following proposition generalizes our result and considers another special case of interest. We begin with two definitions.[30]

Definition 1 (Growth Acceleration Costs) *A more general specification of the GAC function* $\psi : \mathbb{R}^2 \to \mathbb{R}$ *is given by* $\psi(\gamma^{[k]}, \gamma^{[n]}) = h \circ \|\begin{pmatrix} a\gamma^{[k]} \\ (1-a)\gamma^{[n]} \end{pmatrix}\|_p$*, where* $\|\|_p$ *is an* ℓ_p *norm,* $\|\binom{x}{y}\|_p = (|x|^p + |y|^p)^{\frac{1}{p}}$ *for* $1 \leq p < \infty$*,* $a \in [0,1]$ *is a constant and* $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$ *is a function with the following properties:* h *is convex, twice differentiable and fulfills* $(h(x) = h'(x) = 0 \Leftrightarrow x = 0)$ *and* $h''(x)x \geq (p-1)h'(x)$*.*

Note that the simplest choice for h that meets the requirements of Definition 1 is the power function with an exponent above one and at least as high as p. In this case, one could write ψ as $\psi(x,y) = (a_1|x|^p + a_2|y|^p)^{\frac{q}{p}}$ for $a_1, a_2 \geq 0$, $a_1 + a_2 > 0$, $q \geq p \geq 1$, and $q > 1$.

Definition 2 (Zero-Profit Relation and Intercepts) *Define* $\bar{\gamma}^{[k]}$ *and* $\bar{\gamma}^{[n]}$ *as the intercepts of the zero profit relation, defined by* (5)*, with the positive* $\gamma^{[k]}$ *and* $\gamma^{[n]}$ *axes. Specifically, we define the zero-profit* function $zpf(\gamma^{[k]}, \gamma^{[n]}) = \frac{\psi_k(\gamma^{[k]},\gamma^{[n]})(1+\gamma^{[k]})+\psi_n(\gamma^{[k]},\gamma^{[n]})(1+\gamma^{[n]})}{1-\psi(\gamma^{[k]},\gamma^{[n]})} - (\alpha+\beta-1)$ *and the zero-profit* relation *as* $zpr = \{(\gamma^{[k]}, \gamma^{[n]}) \in \mathbb{R}^2 | zpf(\gamma^{[k]}, \gamma^{[n]}) = 0\}$*. The intercepts are then*

$$\bar{\gamma}^{[k]} = \min(\{\gamma^{[k]} \geq 0 | zpf(\gamma^{[k]}, 0) = 0\} \cup \{\infty\})$$
$$\bar{\gamma}^{[n]} = \min(\{\gamma^{[n]} \geq 0 | zpf(0, \gamma^{[n]}) = 0\} \cup \{\infty\}).$$

Lemma 3 (Properties of the Zero-Profit Relation) *In the first quadrant, the zero-profit relation* zpr *describes a single, continuous line that is differentiable and weakly downward-sloping.*

1. *For* $a = 1$*, it is given by* $\gamma^{[k]} = \bar{\gamma}^{[k]}$*.*

2. *For* $a = 0$*, it is given by* $\gamma^{[n]} = \bar{\gamma}^{[n]}$*.*

for short period lengths and thus smaller rates of factor input growth and a small share of the costs ψ in output.

[30] We will focus our attention on the case $\gamma^{[k]} \geq 0$ and $\gamma^{[n]} \geq 0$.

3. *For $0 < a < 1$ both intercepts are finite. Moreover, for $p = 1$ in the continuous-time limit, zpr is linear, $zpr = \{(\xi\bar{\gamma}^{[k]}, (1-\xi)\bar{\gamma}^{[n]}) | 0 \leq \xi \leq 1\}$.*

Proof. In appendix D. ■

Lemma 3 describes some properties of the graph defined by the zero-profit condition (5) for $\gamma^{[k]}, \gamma^{[n]} \geq 0$. Figure 4 illustrates this graph for example specifications of ψ. If GACs only depend on one factor input, the zero profit line is either horizontal or vertical, as in scenarios (a) and (b) in figure 4. The continuity and negative slope of the zero-profit relation guarantees that cost functions that depend on both factors and have the same intercepts are bounded above by these lines.[31] In the continuous-time limit, the ℓ_1 norm generates a linear zero-profit relation, line (c) in figure 4. For ℓ_p norms with $p \geq 2$ in the symmetric case $a = \frac{1}{2}$, the *zpr* curve lies above this linear connection of the intercepts, as in examples (d) for $p = 2$ and (e) for $p = 8$ in the figure. Scenario (f) shows that the *zpr* can in fact lie below line (c), for the example of $p = 1.1$, $q = 3$. For both p and q near 1, this line can be arbitrarily close to the origin. The proof of lemma 3 also shows how the properties of the norm guarantee that the area between the *zpr* and the axes is always star-shaped at the origin.

Proposition 2 (Growth Properties) *Suppose that the aggregate capital-labour ratio grows at a rate $\gamma^{[K]} - \gamma^{[N]} \in [-\bar{\gamma}^{[n]}, \bar{\gamma}^{[k]}]$. For ψ as in Definition 1, the aggregate production sector as described above can be characterized as follows.*

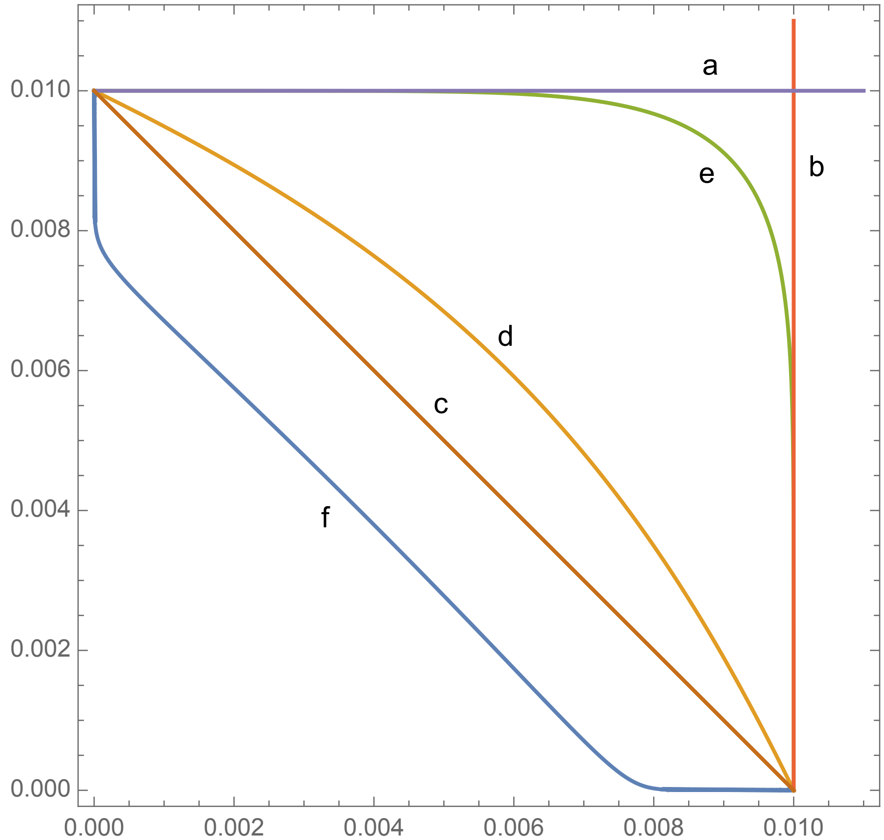
1. *The general case, where both factors are subject to GACs, $0 < a < 1$, at any point in time the following statements are true:*

    (a) *The capital income share lies between $(1-\beta)$ and α.*

    (b) *Both capital and labour input weakly grow at the project level; the rate of total factor productivity growth $\gamma^{[A]}$ at the aggregate level calculated for any Cobb-Douglas production function with a capital share between $(1-\beta)$ and α is bounded,*

[31] All other zero-profit relations (c) to (f) are to the bottom left of (a) and (b) in figure 4.

Figure 4: Zero-Profit Condition for Different GAC Functions ψ



Note: $\psi(x,y) = (a_1 x^p + a_2 y^p)^{\frac{q}{p}}$ in all scenarios; costs either symmetric, $a_1 = a_2$, or only apply to one factor, $a_1 a_2 = 0$. Coefficients a_1, a_2 chosen to keep finite intercepts the same across scenarios, $\bar{\gamma}^{[k]}, \bar{\gamma}^{[n]} \in \{0.01, \infty\}$. Scenarios: (a) $a_2 = 0$, (b) $a_1 = 0$, (c) $p = 1$, continuous time limit, (d) $p = 2$, $q = 3$, (e) $p = q = 8$, (f) $p = 1.1$, $q = 3$.

$-\gamma^{[\psi]} \leq \gamma^{[A]} \leq (\alpha + \beta - 1)\max\{\bar{\gamma}^{[k]}, \bar{\gamma}^{[n]}\} + \gamma^{[\psi]}$, *where* $\gamma^{[\psi]} \in [0, \frac{\psi(\bar{\gamma}^{[k]},\bar{\gamma}^{[n]})}{1-\psi(\bar{\gamma}^{[k]},\bar{\gamma}^{[n]})}]$ *is nonzero only if* $\gamma^{[k]}$ *and* $\gamma^{[n]}$ *fluctuate.*

(c) *The model is consistent with balanced growth in the sense that there is a constant growth rate of the aggregate capital-labour ratio* $\gamma^{[K]} - \gamma^{[N]} \in [-\bar{\gamma}^{[n]}, \bar{\gamma}^{[k]}]$ *for which capital grows at the same rate as output and the interest rate is constant. On a balanced growth path, productivity growth and factor shares are constant, and* $\gamma^{[\psi]} = 0$ *so that TFP growth is nonnegative.*

2. *If only one factor* $i \in \{k, n\}$ *is subject to GACs,* $a \in \{0, 1\}$*:*

(a) *The capital income share is constant,* $(1-\beta)$ *for* $i=k$ *or* α *for* $i=n$*.*

(b) *The corresponding factor input grows at a constant rate,* $\gamma^{[i]} = \bar{\gamma}^{[i]}$*, and total factor productivity growth calculated from the Cobb-Douglas aggregate production function consistent with the observed factor shares is constant, too,* $\gamma^{[A]} = (\alpha + \beta - 1)\bar{\gamma}^{[i]}$*.*

3. *For* $p=1$*, where* $\|\|_p$ *is the* ℓ_1*-norm, in the continuous-time limit:*

(a) *The capital income share is constant at the level* $a(1-\beta)+(1-a)\alpha$*.*

(b) *Total factor productivity growth calculated from the Cobb-Douglas aggregate production function consistent with the observed factor shares is constant, too,* $\gamma^{[A]} = (\alpha+\beta-1)(a\bar{\gamma}^{[k]} + (1-a)\bar{\gamma}^{[n]})$*.*

Proof. In appendix E. ■

As long as the aggregate capital-labour ratio does not change dramatically, so that individual projects can still grow both factor inputs, the economy is rather well-behaved. TFP growth is delimited within a typically very narrow band, and so are the factor income shares. Under appropriate assumptions about the household sector, the economy converges to a balanced growth path, along which capital grows in proportion to labour productivity. On such a path, TFP growth and factor shares are constant. In the special cases described in parts 2 and 3 of proposition 2, growth rates and factor shares are stable even while adjustment dynamics are happening.

Our model fully explains how and at what rate the economy grows, and all resources devoted to improving productivity are fully accounted for, either in the form of added capital at the project level, which includes the fruits of R&D investment[32], or in the form of GACs ψ. In contrast to many existing models of endogenous growth, our firm sector exhibits no scale effects. Neither the level of aggregate employment N nor its growth rate affects the rate of productivity growth on a balanced growth path. Moreover, the model still exhibits full

[32]See appendix A for a version of the model that explicitly accounts for capitalized R&D investments.

adjustment dynamics in the case of deviations of the capital stock from its steady-state level. In alternative endogenous growth models that assume an aggregate production function that is linear in capital, any shocks to the capital intensity typically lead to permanent productivity changes.

2.4 R&D Expenditure

Conceptually, the model splits R&D expenditure into two components. Any expenditure that is appropriable in the sense that it contributes to the value of the firm is an investment that adds to the assets associated with a project, whereas anything else is a cost. This is just a straightforward application of accounting fundamentals to technology and ideas. Appendix A shows that investment in such assets, whether tangible[33] or intangible[34], follows the same principles and mechanisms as investment in "traditional" capital, and we have therefore included these assets in a broader measure of capital, while noting that they contribute to a higher scale elasticity of the firm's production function. This capital accounts for the value of TFP improvements highlighted in popular growth frameworks, such as the value of firms created as a result of R&D efforts in both the Schumpeterian and the varieties models, which are given by the expected present values of profit flows.

What is essential to the core mechanism of our growth framework is a component of R&D expenditure that is directly related to the rate of productivity improvement. We used the symbol ψ for this type of expenditure that is related to the change in production techniques and referred to it as growth acceleration costs. While formally being very similar to the adjustment costs familiar from the investment literature, ψ should be thought of primarily comprising of additional resources required to develop and implement previously unexplored productivity improvements within a set timeframe.

It is the interaction of R&D investment with R&D costs that determines the growth rate.

[33] Any technological knowledge embedded in a tangible asset would typically, for accounting purposes, be considered to contribute to its value.

[34] Including patents, copyrighted works, trademarks, among others.

The availability of opportunities to invest in productivity allows for the creation of assets through R&D, the existence of which contributes to higher scale elasticities. Returns to scale in excess of unity are offset by higher marginal R&D costs, so that the firm statically operates at constant returns to scale.

Empirically, broader measures of R&D intensity, such as reported R&D expenditure or the number of employees engaged in product development, will capture both the capital formation and the cost components of R&D. Other measures such as the value of certain intangible assets or patent counts only include or proxy for the capitalized value of R&D expenditure.

So far, R&D costs have been exactly equal to GACs ψ. The following subsection will show that this was the consequence of our assumptions of free entry combined with any process improvements becoming public knowledge in the following period, and that under alternative assumptions regarding the appropriability of ψ, its cost component can be substantially smaller.

2.5 Market Structure, Competition and Appropriability

The model has been built around assumptions that are consistent with a high degree of competition by assuming both perfectly free entry and exit and price-taking behaviour in all markets. We now investigate the robustness of the core mechanisms with regards to the competitive structure under which firms operate their projects.

There are at least three relevant dimensions to the market structure. First, firms could have static price setting power. Within any given period, they may effectively face demand for their output that presents them with a trade-off between price and quantity. Settings in which this may be the case include monopolistic competition, as well as monopoly or oligopoly at the current point in time. We will incorporate the notion of a downward-sloping demand curve by allowing the firm's revenue $\mathcal{R}_t(y)$ in period t as a function of the quantity sold y to deviate from the linear case $\mathcal{R}_t(y) = y$ that attains under perfect competition.

The second dimension is the protection of firms' innovations with regards to production techniques. Without any protection – the scenario we have worked with so far – the expenditure on ψ cannot contribute to a firm's value beyond the current period, as improvements of the production technique become public knowledge, can be used by any competing firm and can be substituted for by other firms' process innovations which equally become available to the public.

Finally, the degree to which firms are permanently sheltered from entry could matter for outcomes. In the absence of free entry, firms would face predictable demand for their output that is independent of their own and their competitors' actions or the extent of actual or possible profits. We will combine the protection of process innovations and the protection from entry by allowing for the possibility that the firm can rely on predictable revenue functions $\mathcal{R}_s$ for future periods $s > t$ while at the same time enjoying full protection of its production innovations, so that their effects can be fully internalized.

Let $P_t(k, n)$ be the value of a firm operating in a single market that has previously been producing using the technique (k, n) and faces the sequence of revenue functions $\mathcal{R}_s$ for $s = t, t + 1, ...$ for as long as it is sheltered from competition. We will assume that with a probability $\mu \in [0, 1]$, this firm will retain its status as a protected from the effects of entry for another period so that its future value can deviate from zero. Current and future factor prices w_s and r_s are exogenous to the firm's planning.

In every period t, the firm makes two choices involving three variables: the production technique for the projects it controls (k_t and n_t) as well as how much output y_t to sell in total, i.e. what mass h_t of projects to use to satisfy demand in this market, $y_t = h_t(1-\psi)f(k_t, n_t)$. The firm's value can be written recursively as

$$\begin{aligned} P_t(k_{t-1}, n_{t-1}) = \max_{h_t, k_t, n_t} \Bigg\{ \mathcal{R}_t \left(h_t(1 - \psi(\frac{k_t}{k_{t-1}} - 1, \frac{n_t}{n_{t-1}} - 1)) f(k_t, n_t) \right) \\ - (r_t + \delta) h_t k_t - w_t h_t n_t + \frac{\mu}{1 + r_t} P_{t+1}(k_t, n_t) \Bigg\}. \qquad (9) \end{aligned}$$

The case $\mathcal{R}(y) = y$, $\mu = 0$ corresponds to the purely competitive scenario considered before. The optimality conditions for h_t, k_t and n_t are, respectively:

$$\mathcal{R}'_t(y_t)y_t = w_t n_t + (r_t + \delta)k_t \tag{10}$$

$$r_t + \delta = \mathcal{R}'_t(y_t)\left[(1-\psi_t)f_{k,t} - \psi_{k,t}\frac{f_t}{k_{t-1}}\right] + \frac{\mu}{1+r_t}\mathcal{R}'_{t+1}(y_{t+1})\frac{h_{t+1}}{h_t}\psi_{k,t+1}\frac{k_{t+1}}{k_t^2}f_{t+1} \tag{11}$$

$$w_t = \mathcal{R}'_t(y_t)\left[(1-\psi_t)f_{n,t} - \psi_{n,t}\frac{f_t}{n_{t-1}}\right] + \frac{\mu}{1+r_t}\mathcal{R}'_{t+1}(y_{t+1})\frac{h_{t+1}}{h_t}\psi_{n,t+1}\frac{n_{t+1}}{n_t^2}f_{t+1} \tag{12}$$

To find the relationship between the growth rates of factor inputs and the scale elasticity of f, we can eliminate the factor prices from equations (10) to (12). For the interesting special case of stable $\gamma^{[k]}$ and $\gamma^{[n]}$, it is given by

$$\frac{\psi_k\frac{k_t}{k_{t-1}} + \psi_n\frac{n_t}{n_{t-1}}}{1-\psi}\left[(1-\mu) + \mu(r_t - \gamma_{t+1}^{[\mathcal{R}'y]})\right] \doteq \alpha + \beta - 1. \tag{13}$$

For $\mu = 0$, i.e. if there is no persistent protection of process innovations or profits from competitors, the result is identical to equation (5). The characteristics of the revenue function and the resulting possibility to set a price different from marginal costs does not matter at all for the growth outcome.

As the firm's ability to retain its market share improves with rising μ, what changes is the share of the marginal R&D expenditure $(1-\psi)^{-1}(\psi_k\frac{k_t}{k_{t-1}} + \psi_n\frac{n_t}{n_{t-1}})$ that counts as a cost. The more sheltered the firm is from competition, the more it can internalize the effects of current R&D spending on future productivity, thus counting it as an investment that contributes to the firm value rather than a cost. In the case of a perfectly and permanently protected firm, only the interest on this expenditure is considered a cost. While for $\mu = 0$, the *immediate* productivity effect of changes in the production techniques must justify the desired R&D expenditure, with the longer planning horizon for $\mu = 1$ it is sufficient to cover the interest on the spending, net of any appreciation of value of the technology improvement resulting from effective future revenue growth $\gamma^{[\mathcal{R}'y]}$.

All this means that the fundamental mechanism regulating project sizes and growth rates is independent of aspects of the market structure we considered here. While a change in the appropriability of R&D expenditures ψ has a quantitative impact for a given specification of ψ, the general result that marginal GACs offset scale elasticities and firms produce at locally constant returns to scale remains valid. This also means that nothing changes from the perspective of accounting for R&D expenditure as either costs or investment in practical terms.[35]

Everything else equal, better appropriablity of R&D expenditures is associated with higher growth.[36] This important insight has been stressed by much of the more recent literature on endogenous growth, and it is a feature of our framework, too. Here, we looked at it in the context of competition and market structure, but this may not be the only dimension of a firm's environment that matters. To what extent the oft-discussed issue of the legal protection of intellectual property contributes to the relevant appropriability measure and what the right trade-off between dynamic benefits and static inefficiency costs of such measures is, remains a highly relevant question.

The specification we used above presumes a demand function that is entirely independent of project size. Another interesting special case is one where individual projects correspond to product varieties and are each associated with their own, less than perfectly elastic, demand function. As it turns out, in a setting with free entry, this type of static price-setting power again has no impact whatsoever on the growth rate resulting in equilibrium compared to the scenario where firms are price takers in the goods market. The relationship between changes in production techniques and scale elasticities is exactly the same as under price-taking, given by equation (5). The simple derivation of this result is left to the interested reader.

[35] The observation that with higher μ, a larger share of ψ is counted as investment rather than cost may have no deeper meaning at all.

[36] Note, however, that this does not generally imply that sheltered monopolies or oligopolies promise higher growth than a more competitive market structure. In the case of a fixed number of market participants, individual revenue growth is tied to market size growth, possibly resulting in reduced project size growth compared to the competitive scenario where ongoing exit is possible. This issue is discussed briefly in section 3.2 below.

2.6 Comparison with and Discussion of Alternative Models

The key novel feature of our model is the structure of the firm sector, where project size adjusts endogenously to ensure instantaneous constant returns to scale considering all inputs, including technological knowledge, and firms always earn market returns on their R&D investments, while with regards long-term technological opportunities, there are increasing returns.

In the popular R&D-based models of endogenous growth developed in the 1990s, such as the varieties model introduced in P. M. Romer (1990) or the Schumpeterian model of Grossman and Helpman (1991) and Aghion and Howitt (1992), firms' returns on innovation investments are often linked to the size of the economy, typically measured in terms of its labour force, and other model parameters such as demand elasticities in a less flexible way. It is then necessary to assume that R&D costs progress along a particular path, exactly counterbalancing the benefits associated with growing market size and revenue. The result is a model economy that exhibits growth rates that are not stable in the presence of fluctuations in population growth, market size or other parameters.

The scale effects of these growth models, possibly being their most problematic property, have been effectively addressed in various ways. One approach was pursued by the literature on so-called semi-endogenous growth, which ultimately linked aggregate growth outcomes to demographics.[37] Possibly the most compelling solution to the issue of scale effects is due to Young (1998) and Dinopoulos and Thompson (1998), who address the problem in a way that is very similar to our approach: A larger scale of the economy encourages the creation of additional product lines, thus enabling the duplication of R&D effort to absorb the added expenditure on goods into a "horizontal" expansion so that the "vertical" expansion or productivity growth remains unaffected. By ensuring a fixed number of product lines per unit of labour, the model completely separates the processes resulting in productivity growth or quality improvements from scale of the economy, thus justifying the practice of studying

[37]See Jones (1995a), Segerstrom (1998), and Jones (2005).

traditional quality ladder models for fixed labour supply without explicitly addressing the role of the scale of the economy.

If we take this orthogonal decomposition of aggregate output into a scale-dependent number of firms of fixed size and a history-dependent productivity per firm as a starting point for constructing an endogenous growth model, much of the model structure immediately follows from the implicit maintained hypothesis and we are bound to arrive at model economy that closely resembles a canonical quality ladder setting. Letting the current level of technology, which is the same across firms, be A, the labour force N and the possibly endogenous but constant employment per firm $\bar{n}$, the firm-level production function is $A\bar{n}$ and aggregate output is given by AN. We are not accounting for physical capital here, but it would be straightforward to replace the labour input with an intermediate good that is produced according to a constant-returns-to-scale technology involving both capital and labour.[38]

With a constant firm size $\bar{n}$, balanced growth requires that proportional improvements in A must be possible for a fixed amount of labour input, the value of which is itself proportional to A.[39] In other words, devoting a given *value* of resources to R&D always buys the same amount of additional technology A, making TFP A linearly isomorphic to a capital good. The need for such a constant cost of productivity improvements appears to be an arbitrary knife-edge condition. In the absence of clear justification in the form of an economic mechanism that implies this as a robust outcome, the requirement that R&D technology be such that it puts the economy precisely onto the narrow edge between explosive growth and no sustained growth at all remains an exceedingly implausible justification for a phenomenon as stable and robust as economic growth. Moreover, this assumption is at odds with existing evidence that the effectiveness of R&D decreases over time, in other words, that the cost of implementing given productivity improvements rises.[40]

The effective production technology being linear in both labour N and TFP A means

[38] This is not commonly done in practice, most quality-ladder models focus entirely on R&D and labour. Aghion (2004) is an example of a paper explicitly addressing the role of capital.

[39] Otherwise, the share of labour each firm devotes to R&D would either grow or shrink.

[40] See Kortum (1997) and Segerstrom (1998).

that from a societal point of view, output is being produced under a scale elasticity of exactly two. As long as firms are operating in a stable environment, this is not necessarily evident, as A simply rises at a constant rate. The linearity of output in a capital good does, however, have important and well-understood implications. To the extent that individual firms can appropriate R&D investments and thus internalize the nature of TFP as a capital good, any changes to their incentives to invest in it will have persistent growth effects rather than the level effects that would arise were the elasticity of output with respect to R&D capital below unity. Even without appropriability, a range of factors affecting the instantaneous cost and benefit considerations firms are facing can affect long-run growth rates.[41]

Moreover, a social planner or regulator always operates under strongly increasing returns and linear capital. This will result in spurious policy recommendations and exaggerated welfare effects if the elasticity of output with respect to R&D capital is in fact lower than assumed. Nonrivalry in the use of a production factor, the knowledge and ideas that constitute A in our case, suggests that its average productivity effect does not inherently diminish if a given amount is put to use at a larger scale; it does not imply linearity of production in either the value of this factor itself or in a combination of the other remaining factor inputs. Giving such high weight to the productive effect of R&D expenditure seems implausible. If it were possible to increase TFP growth proportionally by increasing or concentrating R&D spending, one might expect that benevolent governments would have found a way of doing so over the past 70 years.

Output being proportional to labour input or a corresponding composite factor implies that an optimizing firm in a competitive environment spends its entire revenue on this input, leaving no room for R&D expenditure. The existence of R&D-based growth might then be taken as evidence that markets cannot be competitive, so that some factors are paid less than their marginal product. Monopolistic competition has been the tool of choice in much of the recent literature on endogenous growth to enable firms to pay factors less than their

[41] What specific policies are growth enhancing always depends on the particulars of the model.

marginal value product and use the mark-up to finance technological improvements.[42] The result that R&D expenditures are closely related to mark-ups then supports the widely held belief that price-setting power of some sort is a prerequisite for growth and, from a dynamic perspective, a desirable feature of markets.

With this we have arrived at the basic structure of a typical quality ladder model. Under the maintained hypothesis stable firm sizes, we have a production technology with significant scale effects, where we treat TFP, despite it having the core characteristics of a capital good, fundamentally different from other inputs, including capital inputs if present. Whereas other inputs are paid according to their marginal value products adjusted for the effects of a firm's price-setting power, R&D expenditure is taken as a residual that is determined largely by a free-entry condition.

One important message of this paper is that, while the structure of the predominant quality ladder framework is very much determined if building on the assumption of a stable size of production units, it is by no means the only structure that can describe a realistic economy where growth results from R&D expenditure.

Relaxing the requirement that firm sizes remain constant allows us to choose the output elasticities of both technology capital and other inputs more freely. For an overall scale elasticity of the atemporal production technology f above unity, including the contribution of knowledge, productivity can still grow at a constant rate as long as the size of production units increases at the right rate. This does not yet ensure that firms can pay factors according to their marginal product given the increasing returns. Our solution to this problem is to assume a cost of moving towards to more productive techniques that is convex in the rate at which this change happens. This idea is being used in quality-ladder settings, for example in models of vertical and horizontal innovation of Klette and Kortum (2004) and Akcigit and Kerr (2018), and seems both fairly general and plausible, highlighting the notion that it is impossible to arbitrarily accelerate R&D without increasing the costs of a given advance, an

[42] In principle, price setting power in the factor markets would open up similar R&D opportunities.

idea that models based on a purely linear R&D technology cannot incorporate. It yields a result that is simple, intuitive and consistent with aggregate evidence. Since instantaneous scale elasticities are now declining with positive rates of growth, the equilibrium growth rate is the one consistent with free entry and zero profits, which by Euler's theorem selects the configuration where returns to scale are locally constant.

We have seen that price setting power in the goods market is not required to support innovation in our model, and in fact, it is entirely neutral. My goal is not to claim that the competitive structure in a market has no effect on growth;[43] however, the chosen model specification allows me to make the point that price setting power is less essential an ingredient in the story of R&D-driven growth than commonly thought while at the same time enabling me to present a simpler, more elegant model.

2.7 Schumpeter Meets Solow

This paper does not propose a radically new way of modelling growth; instead, it builds on the popular quality ladder framework, which is at the heart of the Schumpeterian model of Aghion and Howitt (1992). The original Schumpeterian growth models made the simplifying assumption that each market is supplied by a single firm that has access to the leading technology. Technological innovations could cause the incumbent in the market being supplanted, leading to leap-frogging. Various extensions have relaxed these strong assumptions, resulting in growth models that are capable of creating rich and realistic industry-level dynamics that are in line with Schumpeter's idea of creative destruction.[44] Growth in our model is based on exactly the same fundamental mechanism as in Schumpeterian models, namely the deliberate and incremental improvement of technologies driven by profit considerations, so that any expenditure on R&D must be justified by expected current or future returns. However, with regards to predicted firm dynamics, our base model as presented

[43] See section 3.2 below for further comments on this issue.

[44] Aghion, Akcigit, and Howitt (2014) and Aghion, Akcigit, and Howitt (2015) provide an overview of the developments in in the area of Schumpeterian growth theory.

here approaches from exactly the opposite end of the spectrum compared to the original work of Aghion and Howitt (1992). By assuming symmetry between firms and predictable returns to R&D investment, we *minimize* turbulence and turnover in the firm sector rather than maximizing it.[45] Existing evidence on the degree to which innovation is associated with creative destruction suggests that the truth lies somewhere in between, likely closer to the less turbulent extreme.[46]. Allowing for some degree of randomness and heterogeneity would introduce a larger extent of creative destruction into our model framework, albeit at the expense of increased complexity.

As discussed in section 2.6, popular quality ladder models continue to build on problematic assumptions regarding the characteristics of the R&D technology and overall scale elasticities. At the same time, the Schumpeterian framework has been popular for being able to explain patterns in firm and industry level datasets. By unifying the quality ladder model with the neoclassical growth model, our framework has the potential of replicating these patterns in micro data while still being consistent with mechanics that appear to govern national economies in the long run and at the aggregate level.

The Schumpeterian growth model is well-suited to qualitatively explain the firm size distribution, firm turnover and the relationship between size, age and exit probability, and recent work has shown that it is possible to match many aspects of industry-level outcomes rather well quantitatively. Acemoglu, Akcigit, et al. (2018) and Akcigit and Kerr (2018) are well-known recent examples of doing this successfully.[47]

Incorporating such features in our model framework only requires some relatively straightforward changes. Suppose we assume that firms' production techniques do not become public knowledge immediately so that firms can appropriate some of the expenditure associated with moving to more productive processes. If R&D success is stochastic, both upon entry

[45]The ongoing termination of projects at the rate τ that allows for the concentration required to achieve productivity improvements in our model is possibly the mildest form of firm-level dynamics that might still be considered creative destruction.

[46]See Garcia-Macia, Hsieh, and Klenow (2019).

[47]Other important contributions include Klette and Kortum (2004), Lentz and Mortensen (2008), and Acemoglu and Cao (2015).

and with regards to the outcomes of ongoing R&D efforts, heterogeneity among projects will result. More successful projects will be able to supply larger quantities at lower costs, thus earning rents and having a stronger incentive to innovate further due to their larger sales. In the presence of some knowledge spillovers (imperfect appropriability of ψ), one would expect to see a relatively stable cross-sectional distribution[48] of project characteristics with the following properties. (1) Larger projects spend more on R&D and are more successful on average, (2) unsuccessful projects that shrink below a certain threshold are terminated, (3) because of the endogenous persistence of success, larger firms are older on average and (4) there may be entry of new projects even in industries where, on average, project concentration increases over time.[49]

Compared to a Schumpeterian framework where different firms compete for technological advantages that allow them to replace an incumbent, such as Aghion and Howitt (1992), Acemoglu and Cao (2015), and Acemoglu, Akcigit, et al. (2018), our approach has some potential advantages. First, "product lines" need not be exogenously given, they emerge endogenously with the number of projects serving a particular market. Second, empirically problematic leap-frogging does not arise, and neither does the need to assume complicated strategic interactions between multiple firms in order to avoid this phenomenon. Third, the different market structure in our model may make it possible to study questions or settings that are not as easy to analyze in alternative models. An arbitrary example is the concurrent use of different technologies to produce similar output, which may have different immediate benefits and future prospects.[50]

[48] The overall mass of projects would still typically change over time, as average output per project need not grow at the same rate as demand for the goods supplied by the industry.

[49] Another aspect that could be added is a notion of monopolistic competition combined with investments that can improve demand for varieties.

[50] An example of this is the popularity of steam and electric vehicles in the early 20^{th} century, before the internal combustion engine became the only dominant engine technology in the car industry.

3 Broader Implications

By offering a different theoretical foundation of economic processes within the firm sector, our model can provide new answers to questions and modelling issues in macroeconomics. This section briefly discusses a few of those.

3.1 Exogenous or Endogenous Growth?

The question of whether growth is "exogenous" or "endogenous" has been discussed in the literature for a long time. This is not so much about the true nature of growth in the real world – ultimately economic growth just as any other economic phenomenon should be explainable and thus "endogenous" in an appropriate model; it is about the issue of whether theoretical arguments and data support models of exogenous growth such as the Solow model or more generally the neoclassical growth model or instead models of endogenous growth, such as the Schumpeterian model or the varieties model.

The delineation between the two model worlds comes mostly from the characteristics of the production technology. While exogenous growth is associated with (near) constant returns to scale, endogenous growth is the case of increasing returns, where it is frequently assumed that production is linear in some form of broadly defined capital. Empirically, the different approaches can thus be distinguished based on the effects of investment, including investment in technology and human capital.[51]

The main message of sections 2.2 and 2.3 is that this differentiation between exogenous and endogenous growth is artificial. We have constructed an endogenous growth model that is indistinguishable form the neoclassical growth model at the aggregate level. We have seen that increasing returns in the long run technology do not preclude instantaneous constant returns at the level of industries or national economies.

[51]Mankiw, D. Romer, and Weil (1992) argue that an "augmented" Solow model of exogenous growth that allows for a broader definition of capital matches the data well. For a more recent reassessment see Bernanke and Gürkaynak (2002), including the comment by D. Romer.

In this sense, the need to choose between "exogenous" and "endogenous" growth disappears. Given the characteristics of our endogenous growth model at the aggregate level, the vast theoretical and empirical success of the neoclassical growth model carries over to the realm of endogenous growth.

3.2 Growth and Competition

Empirical studies hint at a possible correlation between the competitive structure in industries and growth outcomes.[52] In the form presented here, our model does not suggest a direct effect of competition on growth. We saw in section 2.5 that price setting power does not per se have implications for growth rates. What matters is the appropriability of R&D expenditure to firms.

It is important to remember that we modelled growth as resulting from *productivity* improvements. The benefit from reducing production costs depends on a firm's sales and should thus be similar for a competitive and a noncompetitive firm of the same size.[53] Generally, all profit maximizing firms will strive to produce efficiently within their given constraints and environments.[54]

One possible place to look for a stronger connection between competition and R&D is in the realm of development of new products and buyer-relevant features. If R&D expenditure enables firms to capture a part of their competitors' market share by offering a more attractive product, the incentive to do so would clearly be stronger in environments where individual market shares are small and competition is fierce.[55] The motivation for such R&D may be similar to the reasons for spending on marketing, although the latter differs from the former with regards to its welfare implications as does not lead to sustained growth.[56]

[52] See for example Nickell (1996) and Blundell, Griffith, and Reenen (1999).

[53] The ability of firms with price-setting power to reoptimize their output in response to a cost change only results in a second-order effect on profits.

[54] Of course the presence of distortionary taxation or inefficient regulation can lead to productive inefficiencies compared to a first-best allocation.

[55] Peters (2020) and Akcigit and Ates (2021) are examples of papers endogenizing firms' market power.

[56] Cavenaile and Roldan-Blanco (2021) is an example of a recently emerging literature on the relationship

Overall, the combination of different incentives to engage in the development of product innovations depending on the nature of the industry and the competitive structure with different degrees of appropriability of R&D expenditures depending on market and product characteristics could likely account for and explain a variety of patterns found in the data.

One particular scenario that is worth mentioning in the context of competition and growth is the case of a natural monopoly. If the market is best supplied by a single project, its growth and scope for productivity improvement will be limited by the market size. If the inability for the project to grow faster than its market imposes an effective constraint, productivity growth will be lower than elsewhere in the economy. Moreover, a lower expenditure on R&D means that marginal GACs do not fully offset returns to scale, so that such industries will have measurably increasing returns to scale and charge prices above marginal costs. Given their slow productivity improvement compared to general wage increases, one would expect them to grow more capital intensive over time.

These considerations may extend to the case where a market is currently served by a small number of projects. Given that large immediate adjustments of the scale of projects is not possible, the exit of one of relatively few players in a market would lead to a drop in supply and likely price increases and rents for the remaining firms. This in turn might deter exit, leading to a situation where a market is supplied by a stable oligopoly of firms that exhibit sub-par productivity growth.[57]

3.3 Profits

When addressing specific questions using macroeconomic models, is often useful to allow for profits in the firm sector. This is particularly true for the current vintage of monetary

between advertising, innovation and firm size characteristics. Dinlersoz and Yorukoglu (2012) and Gourio and Rudanko (2014) look at the role of advertising more generally.

[57]Notice that this view to a certain extent endogenizes the measurable scale characteristics of production technologies in monopolistic or oligopolistic markets. Instead of inferring from observed scale effects in production that the market will and should be served by a few firms or even a single firm, we would argue that only once the productive structure in the market has reached a configuration with so few participants that productivity growth is effectively constrained by market growth, we start measuring scale elasticities above one.

models[58], but applies in other areas as well[59].

The two most popular approaches to implement profits from a modelling perspective are to allow for market power, often in the form monopolistic competition, or to assume diminishing returns to scale. Both approaches have advantages and disadvantages.

Taking our model of the firm sector as a starting point, it is natural to enable firms to make (current period) profits by introducing entry or exit costs.[60] Such costs or barriers have been a common model ingredient in the field of industrial organization, but are not as frequently used in macroeconomics. They would, however, be an inherently plausible source of profits and have several desirable features. They should be fairly easy to calibrate, allow for rich dynamics in response to shocks and make it possible to endogenize market shares and project or firm sizes.

While there would be no market power in such a model, and firms would thus not be able to set prices, they would have entry or exit decisions as an additional strategy variable. Even though these decisions are binary for an individual firm, they result in a continuous market share adjustment when a sufficiently large market is considered.[61]

3.4 Aggregation

With regards to the mechanism underlying the growth model of section 2, a particularly interesting question is: To what extent does the mechanism of endogenous concentration and growth discussed here generalize to settings with heterogeneity among producers? Can an economy, in which a large number of goods and services is produced using different production functions involving different scale elasticities and contributions of capital, still grow at a stable rate and be characterized, at the aggregate level, by a constant-returns-

[58] See Woodford (2003) and Galí (2015) for introductions.

[59] Examples include Blanchard and Kiyotaki (1987), Hopenhayn and Rogerson (1993), and Midrigan and Xu (2014).

[60] While exit costs, looking at the problem statically, appears to have the immediate effect of *reducing* the profits of the firms active in a sector by discouraging exit, the general equilibrium effect is that they discourage entry.

[61] Note that while firms would earn period profits while active in a market, there would still be no expected economic profits, as any excess returns would only offset past or future entry or exit costs.

to-scale technology with constant factor shares? This could give insights into how much heterogeneity among firms or industries matters for growth outcomes.[62,63] Lagos (2006) addresses a similar question for a micro structure based around a model of two-sided search in the labour market.

4 Summary

This paper presented a model that aims to bridge the gap between the world of endogenous growth and the neoclassical growth model. Building on the fundamental ideas incorporated in current theories of endogenous growth, and in particular the quality ladder framework, we have developed a macroeconomic model that accounts for growth-enhancing R&D performed by the firm sector and can explain the sources of ongoing productivity enhancements and the stability of the growth rate while still being consistent with empirically relevant characteristics of aggregate economies, such as constant returns to scale and the existence of largely history-independent steady states with the corresponding adjustment dynamics of capital intensities.

There are two key ingredients that make these results possible. First, the assumption that the rate of productivity improvements is a concave function of R&D expenditure, or equivalently, that the resources required to achieve a given relative increase in output rise superlinearly with output at any point in time, ensures that technological progress is gradual. Second, free entry and exit in industries disconnect the size of production units from market size, thus allowing for stable growth without scale effects or the need for knife-edge calibration of parameters.

[62] A slightly more general version of this question is: Under what conditions can we expect growth to be Harrod neutral or Hicks neutral (Uzawa (1961), Hicks (1932), Robinson (1938), and Harrod (1948))? Moreover, it may be of interest to understand how growth rates and factor shares are affected by disturbances that move the economy off a smooth growth path.

[63] As the growth rates of productivity depend on scale elasticities and *marginal* GACs, whereas the share of R&D expenditure in output is a function of *absolute* GACs as well as the share of investment that would be classified as R&D, a version of this model with heterogeneous firm sizes would not per se make a statement on the Schumpeterian Hypothesis (Schumpeter (1942)).

Under the hood, there are interesting mechanisms at work at the level of firms, markets and industries. From an atemporal perspective, firms are operating under increasing returns to scale. The ever higher productivities reached as time progresses can be thought of as the consequence of innovations that have the character assets; productivity enhancing ideas may take the form of intangible assets or be embedded in better, more valuable capital goods. Either way, they contribute to output and are accounted for in the production function f in the form of a higher capital coefficient, resulting in an overall scale elasticity in excess of unity. Dynamically, the choice of production techniques is limited by the aforementioned requirement to expend resources to move towards improved productivity, an idea we formalized by introducing the growth acceleration cost ψ, which maps a desired change in the production technique to the R&D expenditure required to implement it within a given time. Combining these two concepts results in an instantaneous production function $(1-\psi)f$ that locally has increasing returns to scale at the current factor input combination. However, the further a firm deviates from this already known and tried technique towards higher productivity ones, the lower the scale elasticity becomes at any point in time. Zero profits in the firm sector, the outcome that is consistent with free entry and competition, attains at exactly a scale elasticity of one. This is what pins down the growth rate in a competitive environment for a given technology.

In this way, the model determines the size of production units, which we referred to as projects, in each market in a history-dependent way. For plausible parametrizations, project size grows faster than aggregate output, leading to ongoing concentration of production in fewer units. Being able to endogenously determine market shares opens the door to studying a range of firm and industry level patterns in the context of growth models. Moreover, it could be a beneficial feature in a variety of settings where contemporary macro models rely on non-monolithic firm sectors but arbitrarily assume fixed masses or market shares of units. In part II of this paper, we will see that our firm-sector model is good starting point for developing a business cycle model.

One of the most important features of our model is one that we have not discussed much so far: its simplicity. Having a model available that is easier to teach and learn[64] can make the concept of endogenous growth accessible to a broader audience and help with the communication of theoretical and empirical results and policy recommendations related to growth. The more parsimonious model, where compared to popular growth frameworks less of the structure is pre-imposed,[65] should be easier to use, extend and integrate into other frameworks, facilitating the study of growth-related questions in relation to a variety of other microeconomic and macroeconomic issues. Finally, the much more obvious connection between our notion of endogenous growth and the widely used concept of the neoclassical growth model reduces the fragmentation of macroeconomic theory and emphasizes the idea that the theoretical tools used to address different questions are merely versions of the same fundamental concepts that highlight different aspects of the same underlying reality.

References

Acemoglu, Daron, Ufuk Akcigit, et al. (Nov. 2018). "Innovation, Reallocation, and Growth". In: *American Economic Review* 108.11, pp. 3450–91. DOI: 10.1257/aer.20130470.

Acemoglu, Daron and Dan Cao (2015). "Innovation by entrants and incumbents". In: *Journal of Economic Theory* 157.C, pp. 255–294. DOI: 10.1016/j.jet.2015.01.001.

[64] A basic version of the model can be derived in just a few lines as seen in section 2.2. All important aspects of the technology are open to graphical analysis with the toolkit familiar from production theory. The medium-term dynamics are the same as those of the neoclassical growth model or the Solow model and can be illustrated with the same graphical and analytical tools. Our model does not rely on advanced concepts such as monopolistic competition, Dixit-Stiglitz preferences, nonlinear aggregation, strategic interaction between firms or multiple sectors; in fact, since all core aspects of our model pertain to the production sector, there is no dependence on particular characteristics of the household sector, so that the model can, for example, be closed with an ad-hoc consumption function like the Solow model.

[65] Our framework does not impute a particular elasticity of output with respect to cumulative R&D effort or any other input in the long term, and it does not rely on a particular market structure or on price-setting power of firms. Ultimately, in a basic version of the model where ψ only constrains a single input, it is sufficient to choose three parameters in order to determine all relevant characteristics of the production sector: the coefficients α and β of capital and labour in the production function f as well as one parameter that specifies the shape of ψ. This is enough to pin down the capital and labour share, the growth rate of productivity and the rate of project terminations. This only adds one parameter compared to the neoclassical growth model, where one would choose a capital share and a rate of TFP growth to specify an equivalent production sector.

Aghion, Philippe (2004). "Growth and Development: A Schumpeterian Approach". In: *Annals of Economics and Finance* 5.1, pp. 1–25.

Aghion, Philippe, Ufuk Akcigit, and Peter Howitt (2014). "What Do We Learn From Schumpeterian Growth Theory?" In: *Handbook of Economic Growth*. 1st ed. Vol. 2. Elsevier. Chap. 0, pp. 515–563.

— (May 2015). "Lessons from Schumpeterian Growth Theory". In: *American Economic Review* 105.5, pp. 94–99. DOI: 10.1257/aer.p20151067.

Aghion, Philippe and Peter Howitt (1992). "A Model of Growth Through Creative Destruction". In: *Econometrica* 60.2, pp. 323–351. DOI: 10.2307/2951599.

Ahmad, Mumtaz, John G. Fernald, and Hashmat Khan (Sept. 2019). *Returns to Scale in U.S. Production, Redux*. Carleton Economic Papers 19-07. Carleton University, Department of Economics.

Akcigit, Ufuk and Sina T. Ates (Jan. 2021). "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory". In: *American Economic Journal: Macroeconomics* 13.1, pp. 257–98. DOI: 10.1257/mac.20180449.

Akcigit, Ufuk and William R. Kerr (2018). "Growth through Heterogeneous Innovations". In: *Journal of Political Economy* 126.4, pp. 1374–1443. DOI: 10.1086/697901.

Basu, Susanto and John G. Fernald (1997). "Returns to Scale in U.S. Production: Estimates and Implications". In: *Journal of Political Economy* 105.2, pp. 249–283. DOI: 10.1086/262073.

Bernanke, Ben S. and Refet S. Gürkaynak (Sept. 2002). "Is Growth Exogenous? Taking Mankiw, Romer, and Weil Seriously". In: *NBER Macroeconomics Annual 2001, Volume 16*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 11–72.

Blanchard, Olivier Jean and Nobuhiro Kiyotaki (1987). "Monopolistic Competition and the Effects of Aggregate Demand". In: *The American Economic Review* 77.4, pp. 647–666.

Blundell, Richard, Rachel Griffith, and John van Reenen (1999). "Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms". In: *Review of Economic Studies* 66.3, pp. 529–554.

Cass, David (1965). "Optimum Growth in an Aggregative Model of Capital Accumulation". In: *The Review of Economic Studies* 32.3, pp. 233–240. DOI: 10.2307/2295827.

Cavenaile, Laurent and Pau Roldan-Blanco (July 2021). "Advertising, Innovation, and Economic Growth". In: *American Economic Journal: Macroeconomics* 13.3, pp. 251–303. DOI: 10.1257/mac.20180461.

Dinlersoz, Emin M. and Mehmet Yorukoglu (Apr. 2012). "Information and Industry Dynamics". In: *American Economic Review* 102.2, pp. 884–913. DOI: 10.1257/aer.102.2.884.

Dinopoulos, Elias and Peter Thompson (1998). "Schumpeterian Growth without Scale Effects". In: *Journal of Economic Growth* 3.4, pp. 313–335. DOI: 10.2307/40215991.

— (1999). "Scale effects in Schumpeterian models of economic growth". In: *Journal of Evolutionary Economics* 9.2, pp. 157–185.

Galí, Jordi (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Economics Books 10495. Princeton University Press.

— (Aug. 2018). "The State of New Keynesian Economics: A Partial Assessment". In: *Journal of Economic Perspectives* 32.3, pp. 87–112. DOI: 10.1257/jep.32.3.87.

Garcia-Macia, Daniel, Chang-Tai Hsieh, and Peter J. Klenow (Sept. 2019). "How Destructive Is Innovation?" In: *Econometrica* 87.5, pp. 1507–1541. DOI: 10.3982/ECTA14930.

Gourio, François and Leena Rudanko (2014). "Customer Capital". In: *The Review of Economic Studies* 81.3 (288), pp. 1102–1136. DOI: 10.2307/43551621.

Grossman, Gene M. and Elhanan Helpman (1991). "Quality Ladders in the Theory of Growth". In: *The Review of Economic Studies* 58.1, pp. 43–61. DOI: 10.2307/2298044.

Harrod, R. F. (1948). *Towards a Dynamic Economics*. Macmillan.

Hicks, J. R. (1932). *The Theory of Wages*. Macmillan. DOI: 10.1007/978-1-349-00189-7.

Hopenhayn, Hugo and Richard Rogerson (1993). "Job Turnover and Policy Evaluation: A General Equilibrium Analysis". In: *Journal of Political Economy* 101.5, pp. 915–938.

Jones, Charles I. (1995a). "R & D-Based Models of Economic Growth". In: *Journal of Political Economy* 103.4, pp. 759–784. DOI: 10.2307/2138581.

— (1995b). "Time Series Tests of Endogenous Growth Models". In: *The Quarterly Journal of Economics* 110.2, pp. 495–525. DOI: 10.2307/2118448.

— (Mar. 2002). "Sources of U.S. Economic Growth in a World of Ideas". In: *American Economic Review* 92.1, pp. 220–239. DOI: 10.1257/000282802760015685.

— (2005). "Chapter 16 - Growth and Ideas". In: ed. by Philippe Aghion and Steven N. Durlauf. Vol. 1. Handbook of Economic Growth. Elsevier, pp. 1063–1111. DOI: 10.1016/S1574-0684(05)01016-6.

— (July 2019). "Paul Romer: Ideas, Nonrivalry, and Endogenous Growth". In: *Scandinavian Journal of Economics* 121.3, pp. 859–883. DOI: 10.1111/sjoe.12370.

Kaldor, Nicholas (1966). "Marginal Productivity and the Macro-Economic Theories of Distribution: Comment on Samuelson and Modigliani". In: *The Review of Economic Studies* 33.4, pp. 309–319. DOI: 10.2307/2974428.

Klette, Tor Jakob and Samuel S. Kortum (Oct. 2004). "Innovating Firms and Aggregate Innovation". In: *Journal of Political Economy* 112.5, pp. 986–1018. DOI: 10.1086/422563.

Koopmans, Tjalling C. (1963). *On the Concept of Optimal Economic Growth.* Cowles Foundation Discussion Papers 163. Cowles Foundation for Research in Economics, Yale University.

Kortum, Samuel S. (1993). "Equilibrium R&D and the Patent–R&D Ratio: U.S. Evidence". In: *The American Economic Review* 83.2, pp. 450–457. DOI: 10.2307/2117707.

— (1997). "Research, Patenting, and Technological Change". In: *Econometrica* 65.6, pp. 1389–1419. DOI: 10.2307/2171741.

Kydland, Finn E. and Edward C. Prescott (1982). "Time to Build and Aggregate Fluctuations". In: *Econometrica* 50.6, pp. 1345–1370. DOI: 10.2307/1913386.

Lagos, Ricardo (2006). “A Model of TFP”. In: *The Review of Economic Studies* 73.4, pp. 983–1007. DOI: 10.2307/4123256.

Lentz, Rasmus and Dale T. Mortensen (2008). “An Empirical Model of Growth through Product Innovation”. In: *Econometrica* 76.6, pp. 1317–1373. DOI: 10.2307/40056508.

Lucas, Robert E. (1967). “Adjustment Costs and the Theory of Supply”. In: *Journal of Political Economy* 75.4, pp. 321–334. DOI: 10.2307/1828594.

Mankiw, N. Gregory (Sept. 1989). “Real Business Cycles: A New Keynesian Perspective”. In: *Journal of Economic Perspectives* 3.3, pp. 79–90. DOI: 10.1257/jep.3.3.79.

Mankiw, N. Gregory, David Romer, and David N. Weil (1992). “A Contribution to the Empirics of Economic Growth”. In: *The Quarterly Journal of Economics* 107.2, pp. 407–437.

Midrigan, Virgiliu and Daniel Yi Xu (Feb. 2014). “Finance and Misallocation: Evidence from Plant-Level Data”. In: *American Economic Review* 104.2, pp. 422–58. DOI: 10.1257/aer.104.2.422.

Nickell, Stephen J. (1996). “Competition and Corporate Performance”. In: *Journal of Political Economy* 104.4, pp. 724–746. DOI: 10.2307/2138883.

Peters, Michael (Sept. 2020). “Heterogeneous Markups, Growth, and Endogenous Misallocation”. In: *Econometrica* 88.5, pp. 2037–2073. DOI: 10.3982/ECTA15565.

Prescott, Edward C. (1986). “Theory ahead of business-cycle measurement”. In: *Carnegie-Rochester Conference Series on Public Policy* 25, pp. 11–44. DOI: 10.1016/0167-2231(86)90035-7.

Ramsey, F. P. (1928). “A Mathematical Theory of Saving”. In: *The Economic Journal* 38.152, pp. 543–559. DOI: 10.2307/2224098.

Robinson, Joan (1938). “The Classification of Inventions”. In: *Review of Economic Studies* 5.2, pp. 139–142.

Romer, Paul M. (1986). “Increasing Returns and Long-Run Growth”. In: *Journal of Political Economy* 94.5, pp. 1002–1037. DOI: 10.2307/1833190.

Romer, Paul M. (1990). "Endogenous Technological Change". In: *Journal of Political Economy* 98.5, S71–S102. DOI: 10.2307/2937632.

Schumpeter, Joseph (1942). *Capitalism, Socialism, and Democracy.* New York: Harper and Row.

Segerstrom, Paul S. (1998). "Endogenous Growth without Scale Effects". In: *The American Economic Review* 88.5, pp. 1290–1310. DOI: 10.2307/116872.

Solow, Robert M. (1956). "A Contribution to the Theory of Economic Growth". In: *The Quarterly Journal of Economics* 70.1, pp. 65–94. DOI: 10.2307/1884513.

— (1957). "Technical Change and the Aggregate Production Function". In: *The Review of Economics and Statistics* 39.3, pp. 312–320. DOI: 10.2307/1926047.

Summers, Lawrence H. (1986). "Some skeptical observations on real business cycle theory". In: *Quarterly Review* 10.Fall, pp. 23–27.

Swan, T. W. (1956). "Economic Growth and Capital Accumulation". In: *Economic Record* 32.2, pp. 334–361. DOI: 10.1111/j.1475-4932.1956.tb00434.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-4932.1956.tb00434.x.

Tobin, James and William C. Brainard (1976). *Asset Markets and the Cost of Capital.* Cowles Foundation Discussion Papers 427. Cowles Foundation for Research in Economics, Yale University.

Uzawa, H. (1961). "Neutral Inventions and the Stability of Growth Equilibrium". In: *The Review of Economic Studies* 28.2, pp. 117–124. DOI: 10.2307/2295709.

Woodford, Michael (2003). *Interest And Prices: Foundations Of A Theory Of Monetary Policy.* Princeton, NJ: Princeton University Press.

Young, Alwyn (1998). "Growth without Scale Effects". In: *Journal of Political Economy* 106.1, pp. 41–63. DOI: 10.1086/250002.

Appendix

A Motivation: Modelling Endogenous TFP Investment

In this section, we will extend a simple neoclassical growth model to include endogenous investment in higher total factor productivity. The main objective is to show that investment in new ideas or intellectual property is, from a modelling perspective, no different from investment in physical capital.

Consider a firm that produces output y from two factor inputs, capital k and labour n, according to the production technology $y_t = D_t k_t^{\alpha} n_t^{\beta}$. As in the model of section 2.2, α and β are positive coefficients and D is total factor productivity. Output is perfectly fungible and used as the numeraire. The firm purchases capital at a price of one per unit and rents labour at the wage rate w. The owners of the firm and whoever else helps finance its operations expect a return r on their assets, which include the firm value. The firm can invest in higher productivity. For an R&D investment $\iota(\underline{D}, \overline{D})$, it can raise TFP from $\underline{D}$ to $\overline{D}$. Physical capital k and technological knowledge D depreciate at the rates $\delta^{[k]}$ and $\delta^{[D]}$, respectively.

Technological knowledge D is private but tradable between firms. Firms can invest in R&D in-house, buy technology elsewhere or sell their own technology. When doing so, an amount of technology that originally cost one unit of output to create can replace one unit worth of R&D investment elsewhere, too.

Firms optimally choose how much to invest in physical capital and technology and how much R&D to conduct in-house or acquire externally. We will assume that labour at the firm level is subject to growth acceleration costs (GACs), analogously to the specification used in section 2.2, and will grow at constant rate $\gamma^{[n]}$.[66] All markets are competitive, and there are no barriers to entry or exit. Firms are symmetric with regards to their employment.

[66]The constant average GAC $\psi(\gamma^{[n]})$ is assumed to be accounted for as part of the firm's average productivity D.

Suppose the typical firm starts period t with capital k_{t-1} and technology D_{t-1}. It trades technology, moving the level of TFP to $\tilde{D}_t$. Then it raises productivity to $D_t \geq \tilde{D}_t$ by investing in R&D at an expense $\iota(\tilde{D}_t, D_t)$. Finally, it adjusts physical capital and labour input to their desired levels k_t and n_t and produces its output. During production, capital and technology are subject to depreciation, resulting in additional costs. The value P of such a firm can be defined recursively as

$$P(D_{t-1}, k_{t-1}, n_{t-1}) = \max_{\tilde{D}_t, D_t, k_t} \Big\{ D_t k_t^{\alpha} n_t^{\beta} - p_t^{[D]} \iota(D_{t-1}, \tilde{D}_t) - \iota(\tilde{D}_t, D_t) - (k_t - k_{t-1}) - w_t n_t \\ - \delta^{[D]} \iota(0, D_t) - \delta^{[k]} k_t + \frac{1}{1+r_t} P(D_t, k_t, n_t) \Big\}, \tag{14}$$

where $p_t^{[D]}$ is the market price at which technology is traded and $n_t = (1+\gamma^{[n]})n_{t-1}$.

We will assume that the cost of a marginal technology improvement starting at D is $d(1+\omega)D^{\omega}$, where $\omega > \frac{\alpha}{1-\alpha}$ is a measure of how fast this marginal cost increases with the level of technology and d is a positive constant. The cost of improving technology by a positive amount from $\underline{D}$ to $\overline{D}$ is thus

$$\iota(\underline{D}, \overline{D}) = \int_{\underline{D}}^{\overline{D}} d(1+\omega) D^{\omega} dD = d[\overline{D}^{1+\omega} - \underline{D}^{1+\omega}].$$

Free entry and exit ensure that the value of the firm never deviates from the market value of its assets. Firms can always enter by buying the required amounts of technology and capital at an expense $p^{[D]}\iota(0, D_{t-1}) + k_{t-1}$ or exit and sell their assets for this amount. We thus have $P(D_{t-1}, k_{t-1}, n_{t-1}) = p^{[D]}\iota(0, D_{t-1}) + k_{t-1}$. Substituting this into the firm's value function (14) does two things. First, it shows that the firm's optimization problem is in fact static. Second, it ensures zero profits in the firm sector, as we will see shortly.

The first-order conditions of the firm's problem can, after some straightforward simplifications, be written as:

$$\tilde{D}_t : p^{[D]} = 1 \tag{15}$$
$$D_t : \frac{\partial y_t}{\partial D_t} = k_t^\alpha n_t^\beta \doteq (r_t + \delta^{[D]}) d(1+\omega) D_t^\omega \tag{16}$$
$$k_t : \frac{\partial y_t}{\partial k_t} = \alpha A_t k_t^{\alpha-1} n_t^\beta \doteq r_t + \delta^{[k]} \tag{17}$$

The market price of technology must be $p^{[D]} = 1$, as in-house R&D investment at a cost of unity is a perfect substitute for buying technology in the market. Optimal investment in physical capital and R&D both require the respective marginal products to compensate the firm for interest and depreciation on the marginal unit.

At this point, it is convenient transform our measure of technology, which so far was defined in terms of its contribution to output, into units of replacement costs, as is the usual accounting practice for capital assets. We introduce the asset $b = dD^{1+\omega}$ by equating it to the market value or replacement cost of technology. With this change, the production function can be written as $y = d^{-\frac{1}{1+\omega}} b^{\frac{1}{1+\omega}} k^\alpha n^\beta$. If marginal R&D costs were constant, $\omega = 0$, output would be linear in b. For positive ω there are diminishing returns to this input. The first-order condition for optimal R&D investment (16) simplifies to $\frac{\partial y_t}{\partial b_t} = \frac{1}{1+\omega} d^{-\frac{1}{1+\omega}} b^{\frac{1}{1+\omega}-1} k^\alpha n^\beta \doteq r_t + \delta^{[D]}$.

Finally, the firm's value function (14) can be rearranged to yield

$$y_t - (r_t + \delta^{[k]}) k_t - (r_t + \delta^{[D]}) b_t - w_t n_t \doteq 0.$$

This equation confirms that profits are zero and determines the wage rate. As a residual, the resulting labour share is $1 - \frac{1}{1+\omega} - \alpha$, which is positive under our assumptions regarding ω.

Deriving aggregates as in equation (7), we get

$$Y = AB^{\frac{1}{1+\omega}} K^\alpha N^{1-\frac{1}{1+\omega}-\alpha}, \tag{18}$$

where B, K and N are aggregate technology capital, physical capital and labour and total factor productivity A grows at the rate $(\frac{1}{1+\omega}+\alpha+\beta-1)\gamma^{[n]}$. The model with endogenous investment in TFP thus behaves just like model with exogenous TFP but an extra type of capital, both at the firm and the aggregate level.

Models with multiple capitals have been studied in the growth literature, with Mankiw, D. Romer, and Weil (1992) being a well-known example.

B Growth Acceleration Costs Tied to Capital

Throughout the paper, we assume that GACs scale with output. A similarly plausible assumption would be that they are tied to capital instead. The cost specification $\psi(\gamma^{[k]},\gamma^{[n]})$ introduced in section 2.1 captures both the case where GOCs depend on the existing capital stock and the case where they apply to new capital investment. The former scenario could be one where changes in the production technique require renewal, replacement or reconfiguration of existing capital, such as retooling a factory or updating a complex software product, so that the cost increases both with amount of the existing capital and the rate of change. The latter scenario, which seems closer to mechanisms emphasized in the literature on R&D, includes the case where the acquisition of new, possibly intangible, investment goods becomes relatively more expensive if it needs to happen in a shorter timeframe. The notion that speeding up the R&D process results in a higher cost of the ideas, products and patents created appears to be an important feature of the R&D process.

This appendix shows that growth characteristics of the firm sector do not change in an important way if GACs are tied to capital rather than output as long as these costs increase sufficiently fast with the growth rate. This justifies using just one specification in the rest of this paper.

To see how alternative cost specifications differ, we start by including both types of GACs in the model. Let the profit associated with a project be

$$\pi = f(k,n) - \psi^{[y]}(\gamma^{[k]},\gamma^{[n]})f(k,n) - \psi^{[k]}(\gamma^{[k]},\gamma^{[n]})k - (r+\delta)k - wn, \tag{19}$$

where $\psi^{[y]}$ and $\psi^{[k]}$ are the two types of GACs.

Eliminating the factor prices from the free-entry condition $\pi = 0$ using the first-order conditions for the optimal choice of capital and labour inputs, just as in section 2.1, yields the following equation linking marginal costs to the scale elasticity of f, the equivalent of equation (5) derived with just one type of cost:

$$\frac{\left(\psi_k^{[y]} + \psi_k^{[k]}\frac{k}{f(k,n)}\right)\frac{k}{\bar{k}} + \left(\psi_n^{[y]} + \psi_n^{[k]}\frac{k}{f(k,n)}\right)\frac{n}{\bar{n}}}{1-\psi^{[y]}} = \alpha + \beta - 1 \tag{20}$$

Evidently, the capital-related cost $\psi^{[k]}$ plays a very similar role to the output-related cost $\psi^{[y]}$. In fact, the only differences are that the effect of the marginal cost associated with $\psi^{[k]}$ is scaled by the capital-output ratio, and that no term related to $\psi^{[k]}$ appears in the denominator of the fraction on the left of equation (20).

In order to guarantee that a solution to (20) exists even if $\psi^{[y]} \equiv 0$, the slope of $\psi^{[k]}$ must be able to attain sufficiently high values. Unbounded first derivatives are sufficient, but not necessary to ensure this is the case.

Whenever the economy is in a steady state or on a balanced growth path with a constant interest rate, the capital-output ratio is constant as well, so that the determination of the growth rates $\gamma^{[k]}$ and $\gamma^{[n]}$ works just as in the case of output-related costs discussed in section 2.3, and the results summarized in proposition 2 carry over to the extent that they pertain to balanced growth paths.

Outside of such a steady-state-like situation, however, fluctuations in the capital-output ratio can be expected to induce changes in the project-level growth rate and thus in measured aggregate TFP growth. This can lead to changes in the adjustment path of the economy. Still, some quick back-of-the-envelope calculations reveal that as long as the first partial derivatives of $\psi^{[k]}$ are themselves convex, the variation in firm-level growth rates one would expect to see are extremely small.

In the absence of any catastrophic events, the largest fluctuations in the capital-output ratio one would expect to observe are maybe on the order of 5%, between the peak and trough of the cycle. To ensure that equation (20) still holds, the combined marginal costs $\psi_k^{[k]}$ and $\psi_n^{[k]}$ need to change by 5% in the opposite direction. If the derivatives of $\psi^{[k]}(\gamma^{[k]}, \gamma^{[n]})$ are convex, this means that the corresponding growth rates will change by less than 5% (not percentage points), maybe for example from 4% to 3.8%. This results in a corresponding change in measured aggregate TFP growth, for example from 2% to 1.9%. Fluctuations of this small magnitude are unlikely to be measurable in a reliable fashion. Forward-looking households would consider them a minor exogenous variation in productivity growth that would result in equally minor changes to the optimal adjustment path towards a steady state.

To make this point more formally and precisely, we will show that up to a linear approximation, if $\psi^{[k]}$ has convex directional derivatives and if growth rates are small, the relative deviation of the project-level growth rate from the value $\bar{\gamma}^{[k]}$ that attains at the steady-state capital-output ratio is bounded above by the relative deviation from this ratio. In line with the approach taken elsewhere in this paper, including proposition 2, we will focus on situations where steady-state project level capital growth is positive, $\bar{\gamma}^{[k]} > 0$.

As discussed in section 2.3 and appendix E, the firm-level capital-labour ratio must reflect the aggregate factor supply ratio for markets to clear, which implies that the project-level factor input growth differential is given by aggregate conditions, $1 + \gamma^{[n]} = \Gamma(1 + \gamma^{[k]})$, where $\Gamma = (1 + \gamma^{[N]})/(1 + \gamma^{[K]}) > 0$ is predetermined. Combining this relationship with equation (20) allows us to solve for the growth rates $\gamma^{[k]}$ and $\gamma^{[n]}$. We can write the current capital-output ratio in terms of previous factor inputs $(\bar{k}, \bar{n})$ and current growth rates,

$$\frac{k}{f(k,n)} = \frac{k}{Dk^{\alpha}n^{\beta}} = \frac{\bar{k}(1+\gamma^{[k]})}{D\left(\bar{k}(1+\gamma^{[k]})\right)^{\alpha}\left(\bar{n}\Gamma(1+\gamma^{[k]})\right)^{\beta}} = \Delta(1+\kappa)\Gamma^{-\beta}(1+\gamma^{[k]})^{1-\alpha-\beta},$$

where $\Delta > 0$ is the steady-state capital-output ratio and κ is the relative deviation of $\bar{k}/f(\bar{k}, \bar{n})$ from Δ.

Finally, given the fixed linear relationship between the factor input growth rates in any period, it makes sense to define $\vec{\psi}^{[i]}(\gamma^{[k]}) := \psi^{[i]}(\gamma^{[k]}, \Gamma(1+\gamma^{[k]})-1)$ for $i \in \{y, k\}$, as writing the problem in terms of the relevant directional derivatives will simplify the notation and the interpretation of results.

With this, we can rewrite equation (20) as follows.

$$\left(\vec{\psi}^{[y]\prime}(\gamma^{[k]}) + \vec{\psi}^{[k]\prime}(\gamma^{[k]})\Delta(1+\kappa)\Gamma^{-\beta}(1+\gamma^{[k]})^{1-\alpha-\beta}\right)(1+\gamma^{[k]}) + \\ (1-\alpha-\beta)(1-\vec{\psi}^{[y]}(\gamma^{[k]})) = 0 \tag{21}$$

We linearize equation (21) with respect to $\gamma^{[k]}$ and κ around the equilibrium that attains for the steady-state capital-output ratio Δ, $(\gamma^{[k]}, \kappa) = (\bar{\gamma}^{[k]}, 0)$ using a first-order Taylor approximation, then use $1 + \gamma^{[k]} \doteq 1$ to get

$$\frac{\gamma^{[k]} - \bar{\gamma}^{[k]}}{\bar{\gamma}^{[k]}} \doteq \Theta\kappa,$$

where

$$\Theta = \frac{1}{\bar{\gamma}^{[k]}} \frac{\Delta\Gamma^{-\beta}\vec{\psi}^{[k]\prime}}{(\alpha+\beta)\vec{\psi}^{[y]\prime} + \vec{\psi}^{[y]\prime\prime} + \Delta\Gamma^{-\beta}\left((2-\alpha-\beta)\vec{\psi}^{[k]\prime} + \vec{\psi}^{[k]\prime\prime}\right)}$$

and all derivatives are evaluated at $\gamma^{[k]} = \bar{\gamma}^{[k]}$.

The convexity of $\vec{\psi}^{[k]\prime}$ ensures with $\vec{\psi}^{[k]\prime}(0) = 0$ that $\bar{\gamma}^{[k]}\vec{\psi}^{[k]\prime\prime} \geq \vec{\psi}^{[k]\prime}$. As all other derivatives are nonnegative, it follows that $0 \leq \Theta \leq 1$.

As long $\psi^{[k]}$ has nonnegative third derivatives, meaning it grows at least as fast as a quadratic function, the effect of fluctuations in factor-input ratios on growth rates remains small. This property also ensures that the first derivatives of $\psi^{[k]}$ are unbounded, which is sufficient to guarantee that equation (20) has a solution, as mentioned above.

C Proof of Proposition 1

Suppose the capital-labour ratio is fixed at κ. For a scale of production s, let labour input be λs and capital input $\kappa\lambda s$, and define $\bar{w} = \kappa\lambda(r+\delta) + \lambda w$ as the unit combined factor price. Define $y(s) = (1 - \psi(\lambda\kappa s/\bar{k} - 1, \lambda s/\bar{n} - 1))f(\lambda\kappa s, \lambda s)$ as the instantaneous output as a function of scale. Then, the firm's profit from a project is $\pi = y(s) - \bar{w}s$.

Combining the firm's first-order condition for the optimal choice of scale s, $y'(s) - \bar{w} = 0$, with the the zero-profit condition $y(s) - \bar{w}s = 0$ by eliminating $\bar{w}$, we arrive at

$$y'(s)\frac{s}{y(s)} = 1.$$

The scale elasticity is one at any zero-profit interior optimum, which proves part 3 of the proposition.

Let $\lambda = (D\kappa^{\alpha})^{\frac{-1}{\alpha+\beta}}$, so that we can write $y(s) = (1-\psi(s))s^{\sigma}$ for $\sigma = \alpha + \beta \in [1, 2)$. Here $\psi(s) := \psi(\lambda\kappa s/\bar{k} - 1, \lambda s/\bar{n} - 1))$ is, by construction, strictly convex and has a minimum at $s = \bar{s} \in \mathbb{R}$. We will show that there is at most one $s > 0$ where the firm can be at a zero-profit optimum.

We look for scales s where the scale elasticity of y is unity, $y'(s)s - y(s) = 0$. This condition can be written as $s^{\sigma}\phi(s) = 0$ for $\phi(s) = (\sigma - 1)(1 - \psi(s)) - \sigma\psi'(s)s$. It is clearly satisfied for $s = 0$. Further solutions may arise for $\phi(s) = 0$, but are only admissible if they are in the set $A = \{s > 0 | \psi(s) \leq 1\}$, which may be empty.

In the $\sigma = 1$ case, $s = \bar{s}$, i.e. $\psi'(s) = 0$, is the only candidate for an $s \in A$ for which $\phi(s)$ can be zero.

For $\sigma > 1$, we note that $\phi(s) > 0$ for $\bar{s} > s \in A$ and that $\phi'(s) = -(2\sigma - 1)\psi'(s) - \sigma\psi''(s)s < 0$ for $s \geq \bar{s}$, so that there can be at most one $s \in A$ such that $\phi(s) = 0$, and this s is greater or equal $\bar{s}$.

Either way, there is no more than one scale $s > 0$ with $\psi(s) \leq 1$ such that the scale elasticity at s is one, which given part 3 of the proposition implies the claim 2.

To facilitate the use of more generic expressions, we will use the notation $\alpha_k = \alpha$, $\alpha_n = \beta$, $n_k = k$, $n_n = n$, $w_k = r + \delta$, $w_n = w$. We will use the indices i and j for capital and labour, $i, j \in \{k, n\}$, $i \neq j$, suppress function arguments and write $\tilde{\psi}_i$, $\tilde{\psi}_{ii}$ and $\tilde{\psi}_{ij}$ instead of $\psi_i \frac{n_i}{\bar{n}_i}$, $\psi_{ii} \frac{n_i^2}{\bar{n}_i^2}$ and $\psi_{ij} \frac{n_i n_j}{\bar{n}_i \bar{n}_j}$.

$$\pi = (1 - \psi) f - \sum_{i \in \{k,n\}} w_i n_i = 0 \tag{22}$$

$$\pi_i = (1 - \psi) f_i - \tilde{\psi}_i \frac{f}{n_i} - w_i = 0 \tag{23}$$

$$\pi_{ii} = (1 - \psi) f_{ii} - 2\tilde{\psi}_i \frac{f_i}{n_i} - \tilde{\psi}_{ii} \frac{f}{n_i^2} \tag{24}$$

$$\pi_{ij} = (1 - \psi) f_{ij} - \tilde{\psi}_j \frac{f_i}{n_j} - \tilde{\psi}_i \frac{f_j}{n_i} - \tilde{\psi}_{ij} \frac{f}{n_i n_j} \tag{25}$$

To prove part 1 of the proposition, we need to show $\pi_{ii} < 0$ as well as $\pi_{ii}\pi_{jj} - \pi_{ij}^2 > 0$, using the zero profit condition (22), the first-order conditions (23) and the appropriate expressions for the second derivatives of the profit function (26) and (27).

Note that we have $f_i = \alpha_i \frac{f}{n_i}$, $f_{ii} = -\alpha_i (1 - \alpha_i) \frac{f}{n_i^2}$ and $f_{ij} = \alpha_i \alpha_j \frac{f}{n_i n_j}$. Substituting this in (26) and (27), then eliminating $\tilde{\psi}_i$ using the first-order conditions (23) and $(1 - \psi)$ using the zero-profit condition (22), the second derivatives of the profit function can be written as:

$$\pi_{ii} = -\frac{f}{n_i^2} \left((1 - \alpha_i) \frac{w_i n_i}{f} + (1 + \alpha_j) \frac{w_j n_j}{f} + \tilde{\psi}_{ii} \right) \tag{26}$$

$$\pi_{ij} = -\frac{f}{n_i n_j} \left(-\alpha_j (1 - \alpha_i) \frac{w_i n_i}{f} - \alpha_i (1 - \alpha_j) \frac{w_j n_j}{f} + \tilde{\psi}_{ij} \right) \tag{27}$$

As $\tilde{\psi}_{ii} \geq 0$ follows from the convexity of ψ, it is already clear that $\pi_{ii} < 0$.

Let $A_{ii} = (1 - \alpha_i) \frac{w_i n_i}{f} + (1 + \alpha_j) \frac{w_j n_j}{f}$ and $A_{ij} = \alpha_j (1 - \alpha_i) \frac{w_i n_i}{f} + \alpha_i (1 - \alpha_j) \frac{w_j n_j}{f}$. Notice that $A_{ii} > A_{ij}$, $A_{jj} > A_{ij}$ and $A_{ij} > 0$. We then have

$$\pi_{ii}\pi_{jj} - \pi_{ij}^2 > 0 \Leftrightarrow$$
$$\left[A_{ii} + \tilde{\psi}_{ii}\right]\left[A_{jj} + \tilde{\psi}_{jj}\right] - \left[-A_{ij} + \tilde{\psi}_{ij}\right]^2 > 0$$

Expanding the left hand side of the inequality yields

$$\begin{aligned}
A_{ii}A_{jj} + A_{ii}\tilde{\psi}_{jj} + A_{jj}\tilde{\psi}_{ii} + \tilde{\psi}_{ii}\tilde{\psi}_{jj} - A_{ij}^2 + 2A_{ij}\tilde{\psi}_{ij} - \tilde{\psi}_{ij}^2 &> \\
A_{ii}\tilde{\psi}_{jj} + A_{jj}\tilde{\psi}_{ii} + \tilde{\psi}_{ii}\tilde{\psi}_{jj} + 2A_{ij}\tilde{\psi}_{ij} - \tilde{\psi}_{ij}^2 &\geq \text{[convexity of } \psi] \\
A_{ii}\tilde{\psi}_{jj} + A_{jj}\tilde{\psi}_{ii} + 2A_{ij}\tilde{\psi}_{ij} &> \\
2A_{ij}\left[\frac{1}{2}(\tilde{\psi}_{ii} + \tilde{\psi}_{jj}) + \tilde{\psi}_{ij}\right] &\geq \text{[geometric vs. arithmetic mean]} \\
2A_{ij}\left[\sqrt{\tilde{\psi}_{ii}\tilde{\psi}_{jj}} - |\tilde{\psi}_{ij}|\right] &\geq \text{[convexity of } \psi] \; 0,
\end{aligned}$$

which concludes the proof of part 1.

Note that these proofs apply to the case $\psi = \psi^{[i]}$ as well, as in the proof of part 2, the properties of $\psi(s)$ remain unaffected, and in the proof of part 1, we only required weak convexity of ψ.

D Proof of Lemma 3

We will be using the notation $n(x, y) = \|\binom{x}{y}\|_p$ and $\vec{n}(\varphi, s) = s\vec{n}(\varphi, 1) = \|s\binom{\cos\varphi}{\sin\varphi}\|_p$ with $s \geq 0$ for the ℓ_p norm. As we focus on the properties of *zpr* in the first quadrant only, we can be sure that the norm will only be evaluated for positive arguments. The first and second derivatives of the norm $n(x,y) = (x^p + y^p)^{\frac{1}{p}}$ are then

$$\begin{array}{ll} n_1(x,y) = \frac{n(x,y)}{x^p+y^p} x^{p-1} & n_2(x,y) = \frac{n(x,y)}{x^p+y^p} y^{p-1} \\ n_{11}(x,y) = (p-1)\frac{n(x,y)}{(x^p+y^p)^2} x^{p-2} y^p & n_{22}(x,y) = (p-1)\frac{n(x,y)}{(x^p+y^p)^2} x^p y^{p-2} \\ n_{12}(x,y) = -(p-1)\frac{n(x,y)}{(x^p+y^p)^2} x^{p-1} y^{p-1} & n_{21} = n_{12} \end{array}$$

We begin by showing that the zero-profit relation *zpr* as given in Definition 2, which is a generalization of the linear zero-profit condition shown in figure 3, is still a single continuous line passing through the first quadrant with a nonpositive slope.

Start with the case where the coefficient a in Definition 1 is strictly between zero and one, so that both factor growth rates matter and $\psi(x,y)$ depends on both arguments. According to Definition 2, the relation *zpr* contains the roots of *zpf*, which are given by

$$\begin{gathered} zpf(\gamma^{[k]}, \gamma^{[n]}) = 0 \Leftrightarrow \\ \psi_k(\gamma^{[k]}, \gamma^{[n]})(1+\gamma^{[k]}) + \psi_n(\gamma^{[k]}, \gamma^{[n]})(1+\gamma^{[n]}) = (1-\psi(\gamma^{[k]}, \gamma^{[n]})(\alpha+\beta-1) \Leftrightarrow \\ h'(n(a\gamma^{[k]}, (1-a)\gamma^{[n]}))n_1(a\gamma^{[k]}, (1-a)\gamma^{[n]})a(1+\gamma^{[k]}) + \\ h'(n(a\gamma^{[k]}, (1-a)\gamma^{[n]}))n_2(a\gamma^{[k]}(1-a)\gamma^{[n]})(1-a)(1+\gamma^{[n]}) = \\ (1-h(n(a\gamma^{[k]}, (1-a)\gamma^{[n]})))(\alpha+\beta-1). \end{gathered} \tag{28}$$

Rewriting this equation in terms of polar coordinates (φ, s), $ax = s\cos\varphi$ and $(1-a)y = s\sin\varphi$, we arrive at

$$\begin{gathered} \frac{h'(\vec{n}(\varphi,s))\vec{n}(\varphi,1)}{\cos^p\varphi + \sin^p\varphi}(a\cos^{p-1}\varphi(1+s\cos\varphi) + (1-a)\sin^{p-1}\varphi(1+s\sin\varphi) + \\ (\alpha+\beta-1)h(s\vec{n}(\cos\varphi, \sin\varphi)) = \alpha+\beta-1. \end{gathered} \tag{29}$$

We pick an angle $\varphi \in [0, \frac{\pi}{2}]$ and verify that equation (29) determines a unique $s > 0$ in the corresponding direction. Notice that for a given φ, (29) can be written as $h'(\Theta_1 s)(\Theta_2 + \Theta_3 s) + (\alpha+\beta-1)h(\Theta_1 s) = \alpha+\beta-1$ for positive constants Θ_1, Θ_2, Θ_3. The properties of h

ensure that the left-hand side of this equation is zero for $s = 0$, monotonically increasing and continuous in s and unbounded. For $\alpha+\beta-1 = 0$, we thus have $s = 0$, and for $\alpha+\beta-1 > 0$ there is a unique positive s that solves the equation. The continuity of $zpf(s\cos\varphi, s\sin\varphi)$ in φ and s further guarantees that the mapping from φ to s is continuous as well. We have thus shown that for $\alpha+\beta-1 > 0$, $zpr \cap \mathbb{R}^2_{\geq 0}$ is given by a single continuous line connecting the strictly positive intercepts $\bar{\gamma}^{[k]}$ and $\bar{\gamma}^{[n]}$ of zpr with the positive axes.

In the special case $a = 0$, we can still find $\bar{\gamma}^{[n]}$ as before for $\varphi = 0$. $\gamma^{[k]}$ does not affect ψ in this case, so that the relevant part of zpr is simply given by the vertical line $(\gamma^{[k]}, \gamma^{[n]}) \in \mathbb{R}_{\geq 0} \times \{\bar{\gamma}^{[n]}\}$. The $\gamma^{[k]}$ axis intercept is defined to be infinite in this case in Definition 2, which is the limit of the intercept for $a \to 0$. Similarly, for $a = 1$, the relation is a horizontal line through $\bar{\gamma}^{[k]}$ and $\bar{\gamma}^{[n]} = \infty$. This implies a (weakly) negative slope of zpr in the cases $a = 0$ and $a = 1$.

The total differential of the equation $zpf(\gamma^{[k]}, \gamma^{[n]}) = 0$ can be written as

$$\begin{aligned}&\left[\psi_{kk}(\gamma^{[k]}, \gamma^{[n]})(1+\gamma^{[k]}) + \psi_{nk}(\gamma^{[k]}, \gamma^{[n]})(1+\gamma^{[n]}) + (\alpha+\beta)\psi_k(\gamma^{[k]}, \gamma^{[n]})\right] d\gamma^{[k]} + \\ &\left[\psi_{nn}(\gamma^{[k]}, \gamma^{[n]})(1+\gamma^{[n]}) + \psi_{kn}(\gamma^{[k]}, \gamma^{[n]})(1+\gamma^{[k]}) + (\alpha+\beta)\psi_n(\gamma^{[k]}, \gamma^{[n]})\right] d\gamma^{[n]} = 0, \end{aligned} \tag{30}$$

where, in analogy to the notation used before, ψ_{ii} stands for the second derivative of $\psi(\gamma^{[k]}, \gamma^{[n]})$ with respect to $\gamma^{[i]}$, $i \in \{k, n\}$. Establishing that both terms in square brackets are nonnegative in the $0 < a < 1$, $\alpha+\beta > 1$ case will be sufficient to show that zpr slopes down. Consider the first of these terms. I will suppress function arguments for readability.

$$\begin{gathered}\psi_{kk} \cdot (1+\gamma^{[k]}) + \psi_{nk} \cdot (1+\gamma^{[n]}) + (\alpha+\beta)\psi_k = \\ (h''(n)n_1^2a^2 + h'(n)n_{11}a^2)(1+\gamma^{[k]}) + (h''(n)n_1n_2a(1-a) + h'(n)n_{21}a(1-a))(1+\gamma^{[n]}) + \\ (\alpha+\beta)h'(n)n_1a\end{gathered}$$

Using the requirement $h''(x)x \geq (p-1)h'(x)$ and the fact that $n > 0$, we substitute $(v-1)\frac{h'(n)}{n}$

for $h''(n)$, with $v \geq p \geq 1$.

$$ah'(n)\left[a\left((v-1)\frac{n_1^2}{n}+n_{11}\right)(1+\gamma^{[k]})+(1-a)\left((v-1)\frac{n_1 n_2}{n}+n_{21}\right)(1+\gamma^{[n]})+(\alpha+\beta)n_1\right]=$$
$$\frac{ah'(n)n}{n^{2p}}\left[a\left((v-1)\left(a\gamma^{[k]}\right)^{2(p-1)}+(p-1)\left(a\gamma^{[k]}\right)^{p-2}\left((1-a)\gamma^{[n]}\right)^{p}\right)(1+\gamma^{[k]})+\right.$$
$$\left.(1-a)\left((v-p)\left(a\gamma^{[k]}\right)^{p-1}\left((1-a)\gamma^{[n]}\right)^{p-1}\right)(1+\gamma^{[n]})+(\alpha+\beta)n^p\left(a\gamma^{[k]}\right)^{p-1}\right]\geq 0$$

The nonnegativity of the second term in equation (30) follows by symmetry, which concludes the proof of the nonpositive slope of *zpr* in the first quadrant. With this, only part 3 of Lemma 3 remains to be proven.

Substituting the expressions for the derivatives of the norm, rewrite equation (28) as

$$h'(n)n\left[\frac{\left(a\gamma^{[k]}\right)^{p-1}}{n^p}a(1+\gamma^{[k]})+\frac{\left((1-a)\gamma^{[n]}\right)^{p-1}}{n^p}(1-a)(1+\gamma^{[n]})\right]=$$
$$(\alpha+\beta-1)(1-h(n)). \tag{31}$$

First consider the case of the ℓ_1 norm, $p=1$, in the continuous-time limit, where $1+\gamma^{[i]}\to 1$. The term in the square brackets in equation (31) then simplifies to $\frac{1}{n}$, so that the condition describing the *zpr* becomes $h'(n)=(\alpha+\beta-1)(1-h(n))$. This condition is, by definition, fulfilled for the intercepts $(\gamma^{[n]},\gamma^{[k]})\in\left\{(\bar{\gamma}^{[n]},0),(0,\bar{\gamma}^{[k]})\right\}$, which yield the same value of $n(a\gamma^{[k]},(1-a)\gamma^{[k]})$ that makes the equation hold. The ℓ_1 norm, being linear in the first quadrant, has the same values for convex combinations of the intercepts. Thus, any such convex combination fulfills the zero-profit condition as well, which proves that *zpr* is linear in this case.

E Proof of Proposition 2

To prove the proposition, we rely on the results from Lemma 3 and the simple ideas discussed in the main text and summarized in figure 3. Any equilibrium must simultaneously satisfy the zero-profit condition and market clearing in the factor markets. Under our assumptions, the zero-profit condition is a downward sloping line in the first quadrant, as shown in Lemma 3. Due to symmetry among projects, market clearing requires that project-level capital-labour ratios are equal to the aggregate ratio. This implies that inputs at the project level grow at the rates $\gamma^{[k]} = (1+\gamma^{[K]})/(1-\tau) - 1$ and $\gamma^{[n]} = (1+\gamma^{[N]})/(1-\tau) - 1$ for an exit rate $\tau < 1$. Eliminating τ from these equations, we have the linear relationship

$$1 + \gamma^{[k]} = (1 + \gamma^{[n]}) \frac{1 + \gamma^{[K]}}{1 + \gamma^{[N]}} \tag{32}$$

or $\gamma^{[k]} \doteq \gamma^{[n]} + \gamma^{[K]} - \gamma^{[N]}$ between $\gamma^{[n]}$ and $\gamma^{[k]}$, which is labelled "aggregate factor input ratio" in figure 3. The requirement $\gamma^{[K]} - \gamma^{[N]} \in [-\bar{\gamma}^{[n]}, \bar{\gamma}^{[k]}]$ ensures that this market-clearing line intersects *zpr* in the first quadrant.

As the results of lemma 3 only pertain to the first quadrant, $\gamma^{[k]}, \gamma^{[k]} \geq 0$, the main task will be to show the uniqueness of the equilibrium by arguing that for a given capital-labour ratio, the *zpr* intersects the corresponding factor market clearing condition only once.

Start with the observation that for a norm $\|\|$, a vector p, a nonzero vector d and a scalar x, we have $\left.\frac{\partial\|p+xd\|}{\partial x}\right|_{x=0} > 0 \Rightarrow \forall y > 0 : \left(\|p + yd\| > \|p\| \wedge \left.\frac{\partial\|p+xd\|}{\partial x}\right|_{x=y} \geq \left.\frac{\partial\|p+xd\|}{\partial x}\right|_{x=0} \right)$ if the derivatives exist.[67] In words: If the norm locally increases in a particular direction, moving further in this direction will only yield higher values of the norm and nondecreasing slopes.

Substitute (32) into the zero-profit condition (28) to obtain

$$\begin{aligned} h'(n) n_1 \cdot a (1 + \gamma^{[n]}) \frac{1 + \gamma^{[K]}}{1 + \gamma^{[N]}} + h'(n) n_2 \cdot (1 - a)(1 + \gamma^{[n]}) &= (1 - h(n))(\alpha + \beta - 1) \Leftrightarrow \\ \frac{h'(n)}{1 + \gamma^{[N]}} (1 + \gamma^{[n]}) \left(n_1 \cdot a (1 + \gamma^{[K]}) + n_2 \cdot (1 - a)(1 + \gamma^{[N]}) \right) &= (1 - h(n))(\alpha + \beta - 1), \end{aligned} \tag{33}$$

where again I omit function arguments to improve readability. Evidently, the derivative of the norm $n(a\gamma^{[k]}, (1-a)\gamma^{[n]})$ in the direction $\begin{pmatrix} a(1+\gamma^{[K]}) \\ (1-a)(1+\gamma^{[N]}) \end{pmatrix}$ must be positive for any

[67]The statement regarding levels follows immediately from the convexity of the unit cycle. If I cross an iso-length line towards higher values of the norm, I will never encounter the same iso-length line again while moving in that direction.

combination of growth rates $\begin{pmatrix} \gamma_0^{[k]} \\ \gamma_0^{[n]} \end{pmatrix}$ that satisfies this equation. Then, along the ray $\begin{pmatrix} a\gamma^{[k]} \\ (1-a)\gamma^{[n]} \end{pmatrix} = \begin{pmatrix} a\gamma_0^{[k]}+a(1+\gamma^{[K]})x \\ (1-a)\gamma_0^{[n]}+(1-a)(1+\gamma^{[N]})x \end{pmatrix}$ for $x \geq 0$, n is higher and the directional derivative is not lower than at $\begin{pmatrix} \gamma_0^{[k]} \\ \gamma_0^{[n]} \end{pmatrix}$. Given that both h and h' increase in n, equation (33) cannot be satisfied for any other point along the ray. As the ray in question is a subset of the line (32), we have shown that for given $\gamma^{[K]}$ and $\gamma^{[N]}$, there is at most one combination of $\gamma^{[k]}$ and $\gamma^{[n]}$ that satisfies both the zero-profit condition and factor market clearing (32).[68]

The rest of the proof is relatively straightforward.

1. General Scenario

   (a) This follows immediately from (3) and (4) with the nonnegativity for ψ_k and ψ_n for $\gamma^{[k]}, \gamma^{[n]} \geq 0$.

   (b) Deriving aggregate output along the lines of (7) by either eliminating k or n, we arrive at the two expressions

$$Y_t = (1-\psi)Dn_t^{\alpha+\beta-1}K_t^{\alpha}N_t^{1-\alpha} = (1-\psi)Dk_t^{\alpha+\beta-1}K_t^{1-\beta}N_t^{\beta}. \tag{34}$$

   The growth accounting equation assuming a capital share of $\alpha x + (1-\beta)(1-x)$ for $x \in [0,1]$ can be obtained as a convex combination of the equations derived from the two expressions for aggregate output.

$$\begin{aligned} \gamma^{[Y]} \doteq & \frac{-\Delta\psi}{1-\psi} + (\alpha+\beta-1)(x\gamma^{[n]} + (1-x)\gamma^{[k]}) + \\ & [\alpha x + (1-\beta)(1-x)]\,\gamma^{[K]} + [(1-\alpha)x + \beta(1-x)]\,\gamma^{[N]} \end{aligned} \tag{35}$$

   The two terms on the right of the equation in the first line represent the change

[68] This approach to showing that for $\gamma^{[k]}, \gamma^{[n]} \geq -1$, the *zpr* is the edge of a set that is star-shaped at $\gamma^{[k]} = \gamma^{[n]} = -1$ is a relatively easy way of showing a global property of *zpr* and may thus be a good starting point for various generalizations of this proposition. Such generalizations could include more general norms and characteristics of ψ, an extension of the results beyond the first quadrant, i.e. $\gamma^{[k]} < 0$ or $\gamma^{[n]} < 0$, and additional production factors, such as different types of capital.

in total factor productivity $\gamma^{[A]}$. Given the bounds $0 \leq \gamma^{[i]} \leq \bar{\gamma}^{[i]}$ for $i \in \{k, n\}$, the second term must be between 0 and $(\alpha + \beta - 1)\max\{\bar{\gamma}^{[k]}, \bar{\gamma}^{[n]}\}$. The first term measures how much the change in the value of ψ directly affects net final output. The bounds used in the proposition result from the consideration that $0 = \psi(0,0) \leq \psi \leq \psi(\bar{\gamma}^{[k]}, \bar{\gamma}^{[n]})$ in any period.

(c) For a constant growth rate of the aggregate capital-labour ratio (clr), one constant combination of $\gamma^{[k]}$ and $\gamma^{[n]}$ fulfills *zpr* and it follows that ψ is constant and $\gamma^{[\psi]} = 0$. On a balanced growth path, clr grows at the same rate as labour productivity (lp). Using (34), we find a difference between the growth rates of clr and lp of $\bar{\gamma}^{[k]} - \left[(\alpha + \beta - 1)\bar{\gamma}^{[k]} + (1-\beta)\bar{\gamma}^{[k]}\right] > 0$ for the highest possible value of $\gamma^{[K]} - \gamma^{[N]} = \bar{\gamma}^{[k]}$ and the corresponding value of *zpr*, $\gamma^{[k]} = \bar{\gamma}^{[k]}$, $\gamma^{[n]} = 0$. Similarly, for $\gamma^{[K]} - \gamma^{[N]} = -\bar{\gamma}^{[n]}$ we get $-\bar{\gamma}^{[n]} - \left[(\alpha + \beta - 1)\bar{\gamma}^{[n]} - \alpha\bar{\gamma}^{[n]}\right] < 0$. Continuity ensures that there is a value of $\gamma^{[K]} - \gamma^{[N]}$ in between these bounds for which clr and lp grow at the same rate. For the resulting $\gamma^{[k]}$ and $\gamma^{[n]}$, according to (3) the interest rate is constant as well. With the constant capital-output ratio, this implies constant factor income shares.

2. These results were already derived in the main text for the case $i = k$. $i = n$ follows by symmetry.

3. ℓ_1 Norm

   We begin by observing that in this case, ψ, ψ_k and ψ_n are all constant. Details on the derivatives of the norm can be found in appendix D.

   (a) Let n be the (constant) value of the norm on *zpr*. We have $\psi_k = ah'(n)$ and $\psi_n = (1-a)h'(n)$. Substituting this into (3) and (4) and using the continuous-time simplification $\bar{n} = n$ and $\bar{k} = k$, we get

$$\frac{(r+\delta)k}{y} = \alpha - ah'(n)$$
$$\frac{wn}{y} = \beta - (1-a)h'(n).$$

Eliminating $h'(n)$ from these equations and setting $y = (r+\delta)k + wn$ yields the result.

(b) The TFP growth rate can be obtained from (35) for $\Delta\phi = 0$ and $x = (1-a)$.