[redacted]

 WP/19/294

IMF Working Paper

Research Department

Innovate to Lead or Innovate to Prevail: When do Monopolistic Rents Induce Growth?

Prepared by Roberto Piazza and Yu Zheng

Authorized for distribution by Helge Berger

December 2019

Abstract

This paper extends the Schumpeterian model of creative destruction by allowing followers' cost of innovation to increase in their technological distance from the leader. This assumption is motivated by the observation the more technologically ad- vanced the leader is, the harder it is for a follower to leapfrog without incurring extra cost for using leader's patented knowledge. Under this R&D cost structure, leaders innovate to increase their technological advantage so that followers will eventually stop innovating, allowing leadership to prevail. A new steady state then emerges featuring both leaders and followers innovating in few industries with low aggregate growth.

JEL Classification Numbers: 031, 034, 041, L16

Keywords: Innovation,growth, creative destruction, R&D cost

Author's E-Mail Address: rpiazza2@imf.org, yu.zheng@qmul.ac.uk

* We gratefully acknowledge helpful comments from: Phillipe Aghion, Michele Boldrin, Russell Cooper, Mario Crucini, Domenico Ferraro, Gino Gancia, Berthold Herrendorf, Yang Lu, Rodolfo Manuelli, Ramon Marimon, Stephen Parente, Ed Prescott, and Michael Song. We also thank seminar and conference par- ticipants at the Arizona State University, Washington University in St. Louis, LSE, PSE, Queen Mary University of London, Chinese University of HK, Birkbeck Centre for Applied Macro Workshop 2019, CfM London Macro Workshop 2019, Southwestern University of Finance and Economics, RES Meetings 2019, HKUST-Jinan Conference on Macroeconomics 2019, SAET Conference 2018, Midwest Macro Meetings 2018. An earlier version of this work was circulated under the title "Leadership Contestability, Monopolistic Rents and Growth" (IMF WP 11/63).

† Research Department, International Monetary Fund. Address: 700 19th Street, N.W., Washington D.C. 20431. Email: rpiazza2@imf.org.

‡ School of Economics and Finance, Queen Mary University of London. Address: Mile End Road, E1 4NS London, UK. Email: yu.zheng@qmul.ac.uk

1 Introduction

The Schumpeterian idea of creative destruction lies at the heart of the innovation-based growth theory (Grossman and Helpman, 1991; Aghion and Howitt, 1992). Under this view, growth is achieved by entrepreneurs or firms successively undertaking costly R&D to improve and replace each other's exiting product. There are two key ingredients to Schumpeter's theory of creative destruction. First, a successful innovator or leader, by launching a new product in the market, reveals the frontier knowledge embodied in the product to potential competitors or followers. The full knowledge spillover ensures a level playing field in the next round of innovation race. Second, to prevent the very knowledge spillover from enabling an imitator in the product market to drive the product price down to marginal cost, monopoly rights must be granted to the leader whereby he can recoup his initial investment in R&D to bring about the product in the first place. The rich dynamics of competition, firm exit and turnover inherent in the Schumpeterian model makes it the bedrock on which a large and growing literature is based, where the outcomes from creative destruction among heterogeneous firms can be mapped to micro data (starting from Klette and Kortum (2004)).

In this paper, we call into question a key assumption of Schumpeterian model and show how some of its main implications get turned on their heads. The assumption is the assertion that full knowledge spillover enables all, leaders and followers alike, to compete equally in the innovation race for the next product. We believe that this assumption is too drastic and unrealistic. We claim that, even when a leader's knowledge is made public, for example through the patent's registration process, it is far from immediate that followers can effectively make a productive use of it. Because of the very nature of the patent system, any innovation that builds upon previously patented knowledge would face costly legal challenges by the industry leader and, in any case, the amount of "costless" knowledge that is revealed through a patent's application is always limited - as any student knows, knowledge is never really for free and learning is a costly, time consuming process.

Examples of the disadvantage that a follower faces trying to improve on a leader's product abound in history and at present. In 1769, the great inventor James Watt obtained a patent on his idea of a separate condenser in a steam engine, an improvement upon the Newcomen steam engine. Over the next thirty years when his patent lasted, steam engines were modified and improved by many of his peers: William Bull, Richard Trevithick, Arthur Woolf, and Jonathan Hornblower. Yet none of these models made it to the market until 1804 after the patent expired. Because no matter how much better the newer models were, they had to use the idea of the separate condenser (Boldrin and Levine (2008) contains many more examples). Fast forward 250 years, in 2007, Apple and Samsung began their decade-long multi-million dollar patent war that spread across courts in ten countries around the globe. If knowledge cannot be put into productive use, it is to followers what water and fruits are to Tantalus.

The discussion above indicates that innovation costs for a follower are at least as large as, if not larger than, those for a leader, and that the technologically more advanced a leader is relative to the follower, the more costly it is for the follower to leapfrog him. At the extreme, when the technological distance is large enough, then the leader has achieved the goal of his endpoint strategy of pushing the innovation race to a state where any attempt by the follower to leapfrog the incumbent has become prohibitively expensive, and innovation efforts by both firms cease. The asymmetry between leader's and follower's innovation capability has long been addressed by the theoretical microeconomic literature on races (tracing back to Harris and Vickers (1987); Budd et al. (1993), and more recently Hörner (2004)). It is exactly the intuition developed in this literature that we embed in our endogenous growth model. This strategic pattern of innovation by leaders is consistent with empirical observations at the industry level. Using Compustat data, we plot the level or the growth rate of R&D by an industry's leader as a function of his distance from the industry's followers, proxied by either his market share (the top panels of Figure 1) or the duration of his leadership (the bottom panels of Figure 1).[1] An inverted-U relation emerges: when a firm holds a clear leadership in an industry (defined, for instance, by a market share above 50% or a leadership established by at least 10 years), its R&D effort falls as its advantage increases.

To state our case in the simplest possible way, we embed the assumption of asymmetric R&D costs into an otherwise standard endogenous growth model à la Grossman and Helpman (1991), where followers' R&D efforts are aimed at leapfrogging the leaders, and thus contribute to aggregate growth. Specifically, we study the effects of state-dependent innovation costs in a general equilibrium model with a continuum of industries, where in each industry a leader and a follower play a game of innovation. The state of the industry is the technological distance between the leader and followers, so that when followers fall behind in the innovation race they see their innovation costs rise. In this model, the balance growth path where only followers innovate (i.e. Grossman and Helpman (1991)) is no longer the only one that can emerge in equilibrium. For certain range of parameters, the high-growth equilibrium, i.e. the Grossman-Helpman-type steady state, can coexist with "growth traps," which are low-growth equilibria where also leaders innovate provided that their technological

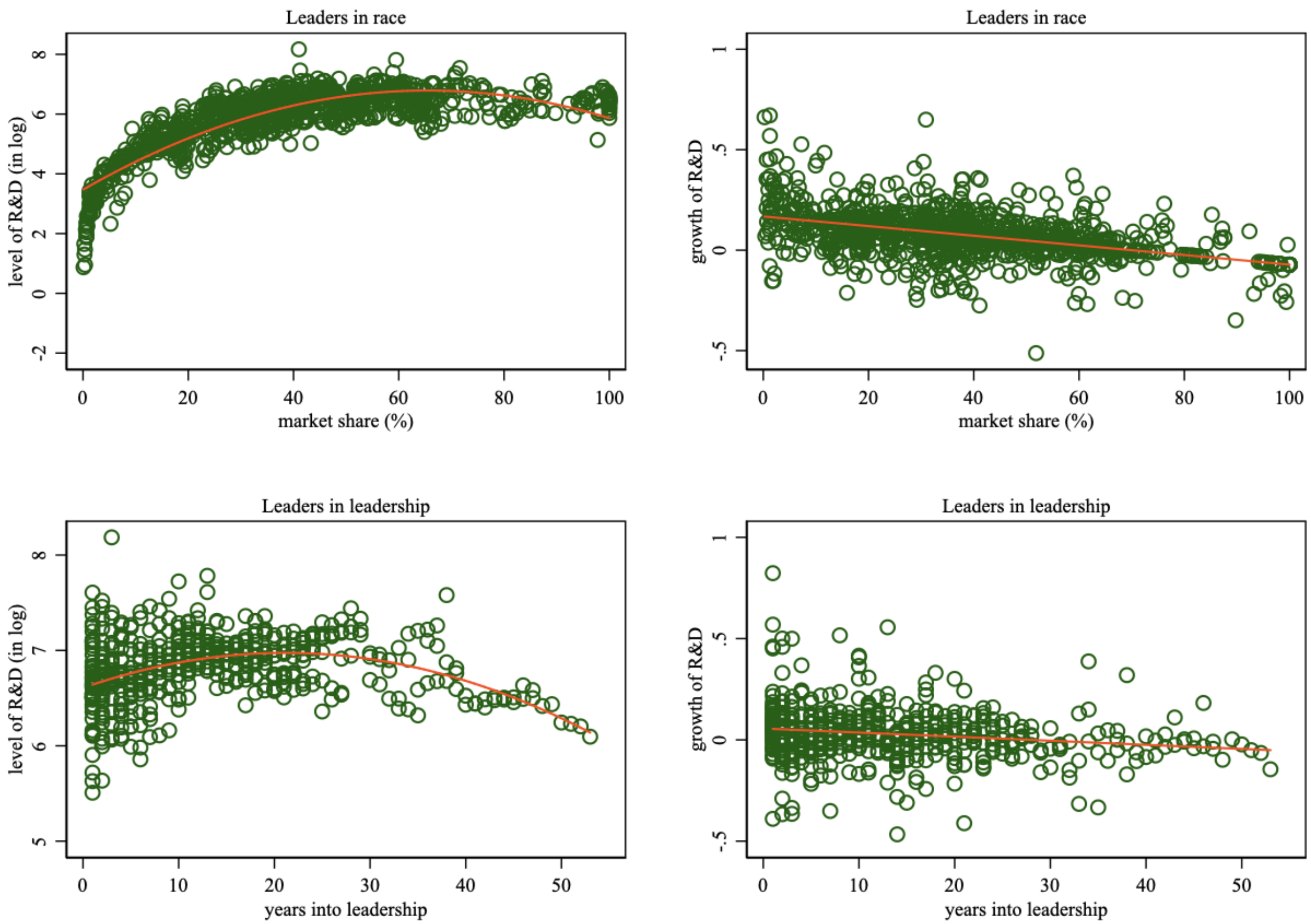
[1]We extract from Compustat a panel of firms that have been an industry leader in any year between 1962 and 2018 in the top 10 R&D intensive industries. For detailed descriptions of the sample construction and the regressions which produce this figure, see Appendix E.

gap with the followers is small. These equilibria are characterized by low growth, because the R&D effort by leaders has two opposite effects on the aggregate innovation rate. On one hand, it contributes to raising the innovation rate (the intensive margin of innovation) of an industry where leader and followers are close to each other. On the other, it increases the share of industries (the extensive margin) where leaders and followers are sufficiently far enough from each other that innovation drops to zero. We show that the positive effect on the intensive margin is always dominated by the negative effect on the extensive margin, so that the success of leaders' end-point strategies is both the reason for large innovation being done at times by leaders, and the cause of low aggregate growth.

In the literature, the asymmetry in R&D costs has been addressed only tangentially. In their seminal work, Klette and Kortum (2004) recognize that leaders possess a "knowledge capital" that increases the productivity of R&D performed by leaders. However, they do not explore how this could affect the innovation race *within* a given industry and assume instead that innovation done by leaders is aimed at entering *new* product lines by leap-frogging the respective incumbents. From the point of view of the strategic interaction between leaders and followers within an industry, their model thus replicates the characteristics of a standard quality ladder as in Grossman and Helpman (1991). The idea that long-lasting monopolies can be detrimental to growth is explored in Aghion et al. (2005). In that paper, the authors argue that the inverted-U pattern can be explained by a market structure that features an "escape competition" effect, built on assumptions (i.e. Bertrand limiting pricing) that guarantee that monopolistic profits are increasing in the distance between leader and follower. We reach the same conclusion by emphasizing, however, a different mechanism, one that is based on the asymmetric R&D cost structure. More recently, Acemoglu and Akcigit (2012) have studied innovation efforts that are dependent on the technological distance. The richest version of their model (i.e. the "leapfrogging and infringement" extension) allows for both step-by-step slow catch-up and "frontier" R&D by followers, the latter of which resembles our cost assumption. In our view, the step-by-step catch up process seems a less realistic description than the costly leapfrogging of the innovation behavior of followers. The presence of patent infringement threats forces followers to find new ways to produce better goods, albeit at a much higher cost, rather than to retrace the leader's footsteps at a lower speed. The main departure of our work from theirs is thus to zoom in on the costly leapfrogging and provide a full analytical characterization of the possible equilibria. In contrast, their study is mainly quantitative, where simulation results are influenced by multiple innovation processes of the followers.

The rest of the paper proceeds as follow. Section 2 introduces the baseline model, characterizes the steady states, and simulates transitional dynamics. Section 3 provides simulations where we vary the policy variables in the model and discuss policy implications. Section 4 discusses two extensions of the baseline model. Conclusion follows.

Figure 1: Share of R&D Conducted by Industry Leaders



Note: The top panels correspond to regressing the level and growth of R&D on market shares over the lifecycle of an eventual leader. The bottom panels correspond to regressing the level and growth of R&D on the age of leadership for leaders. The green circles are firm-year observations after removing the fixed effects from the regressions. The red curves provide the fitted values of the regression, not including the fixed effects. See Appendix E.

2 The Baseline Model

The model is based on the seminal model of quality ladders in Grossman and Helpman (1991) (simply GH from now on). It is a continuous time infinite horizon model. There is a continuum of goods, indexed real numbers in a unit interval. There are two types of agents in the model, households and firms.

2.1 Households

There is a representative household who decides what to consume at each point in time, given its income. It is endowed with one unit of labor and supplies it inelastically. It owns the firms in the economy and hence receives a stream of profits from the firms. Its wealth at time 0, $W(0)$, is then the present value of the stream of profits and labor income it receives ad infinitum. At each instant, the household chooses the quantity, $d_t(i)$, of each of the $i \in [0,1]$ goods to consume, taking as given the quality of each good, $q_t(i)$, the price of each good, $p_t(i)$, and the instantaneous interest rate $r(t)$.

The household consumes $C(t)$ at time t, which is an aggregate of all varieties of goods:

$$\log C(t) = \int_{[0,1]} \log\left(q_t(i) d_t(i)\right) di. \tag{1}$$

The functions $q_t(i) > 0$ define the highest quality developed up to time t for good i. The household's lifetime utility is characterized by a time-additive log period utility function with a rate of time preference of ρ. It solves the following problem:

$$\max_{\{d_t(i),\forall i\}_{t=0}^{\infty}} \quad \int_0^\infty e^{-\rho t} \log C(t) dt \tag{2}$$

$$\text{s.t.} \quad \int_0^\infty e^{-R(t)} E(t) dt \leq W(0), \tag{3}$$

where $R(t)$ is the compounded interest rate and $E(t)$ represents total spending at time t:

$$R(t) = \int_0^t r(\tau) d\tau,$$

$$E(t) = \int_{[0,1]} p_t(i) d_t(i) di.$$

The Cobb-Douglas form of the the consumption aggregate implies that the amount spent by the household on good i is the same across all products, giving

$$d_t(i) = \frac{E(t)}{p_t(i)}.$$

The intertemporal Euler equation gives

$$\frac{\dot{E}(t)}{E(t)} + \rho = r(t). \tag{4}$$

Household's wealth $W(0)$ is given by

$$W(0) = \int_0^\infty e^{-R(t)} \left[\Pi(t) + w(t)L(t) + w(t) \int_{i \in [0,1]} \omega_t(i)\Lambda_t(i)]di \right] dt,$$

where $\Pi(t)$ are aggregate profits received from firms, $w(t)$ is the wage paid to labor employed in the production sector, $L(t)$, and $\omega_t(i)$ is the wage premium paid to (skilled) labor employed in the R&D sector in industry i, $\Lambda_t(i)$. We refer to $\Lambda(i)$ as the *intensive margin* of innovation in industry i. The role of these variables is explained in detail in Section 2.2, which lays out firms' problem. Here we simply specify that total labor is in fixed supply, normalized to unity

$$L(t) + \int_{i \in [0,1]} \Lambda_t(i)di = 1,$$

where $L(t)$ and $\Lambda_t(i)$ are all non-negative. We also assume that the intensive margin of innovation must be bounded above by some constant $\bar{\Lambda}$. The interpretation is simply that there is at most an amount $\bar{\Lambda}$ of workers in the economy with the necessary skill to perform R&D activities in any given industry. For example, there is a fixed supply of labor skilled in biomedical sciences available to the pharmaceutical industry, a fixed supply of labor skilled in computer science available to the information technology industry, so on and so forth. Clearly, in this situation the household's optimal supply of skilled labor to R&D in an industry, $\Lambda^*(i)$, is the correspondence $\Lambda^*(i) = [0, \bar{\Lambda}]$ if the wage premium is equal to one, while $\Lambda^*(i) = \bar{\Lambda}$ whenever $\omega(i) > 1$. Modeling the supply of skilled labor as perfectly elastic up to $\overline{\Lambda}$ and perfectly inelastic afterwards has two advantages. First, when $\overline{\Lambda}$ does not bind, our model is equivalent to GH's model. In other words, ours nests GH. Second, this is a simple and intuitive way to introduce decreasing returns to skills at the aggregate level into the model. We take $\overline{\Lambda}$ as an arbitrarily large constant.[2]

2.2 Firms

Each product i corresponds to an industry. In each industry, there is a leader and a competitive fringe of followers. The leader in industry i has the technology to produce the state-of-the-art version $q_t(i)$ of product i. Such technology is protected by a patent, so that only the leader can produce the quality $q_t(i)$. Leaders and followers also carry out R&D activities. A successful innovation by either a leader or a follower raises the state-of-the-art quality from $q_t(i)$ to $\gamma q_t(i)$, where $\gamma > 0$ is the distance between two consecutive rungs on product i's quality ladder. The quality of a good can then be written as $q_t(i) = \gamma^{s_t(i)}$, with $s_t(i) \in \mathbb{N}$ the number of rungs along the quality ladder that have been climbed up to time t in industry i. Detailed description of firms' production and R&D activities follow.

2.2.1 Production of Goods

The output $y_t(i)$ in industry i is produced using labor $l_t(i)$ according to a linear technology

$$y_t(i) = l_t(i).$$

Since for a given industry, products of different qualities are perfect substitutes, a leader who charges a markup over marginal cost of production labor anywhere between 1 and γ can put the followers, who are at least one rung behind him, out of business. Let the markup charged by leaders be $m \in [1, \gamma]$, which we interpret as a policy variable exogenously determined, as when, for instance, an antitrust authority limits the monopoly pricing power of the leaders. We will later investigate how the growth rate of the economy varies with the markup level m. The price at which leaders sell their products is therefore given by

$$p_t(i) = mw(t), \text{for } m \in [1, \gamma], \tag{5}$$

[2]For a large enough $\overline{\Lambda}$, we obtain the richest implications from the model with three steady states, which we can rank by the respective steady state growth rate. More specifically, the value of $\overline{\Lambda}$ determines the growth rate of the economy in the low growth steady state, as will be explained later. On a more technical note, to obtain an equilibrium under such a supply function with a kink, we proceed in two steps. First, we propose a continuously differentiable supply function with a parameter that governs the speed the supply increases as λ exceeds $\overline{\Lambda}$. Second, we let the parameter go to infinity so at the limit we obtain the desired supply function. Details can be found in Appendix A.

where $w(t)$ is the wage rate of production labor. Without loss of generality, we normalize $w(t)$ to 1 so that, from now on, we express variables in terms of the period wage. The goods prices are then $p_t(i) = m$ and profits of leaders can then be simply expressed as

$$\pi_t(i) = (m-1)y_t(i).$$

At equilibrium prices, the household's demand for good i is given by

$$d_t^*(i) = \frac{E^*(t)}{m}.$$

Using the market clearing condition, $d_t^*(i) = y_t^*(i)$, we conclude that all industries produce the same amount of output, $Y(t)$, using the same amount of labor, $L(t)$, given by

$$Y^*(t) = L^*(t) = \frac{E^*(t)}{m}. \quad (6)$$

It follows then that, for all leaders, profits are given by

$$\Pi^*(t) = (m-1)L^*(t). \quad (7)$$

2.2.2 Game of Innovation

Within each industry, leaders and followers play a game of innovation, and expectations about each other's future strategies determine current innovation efforts. Let's start by describing the innovation technologies available to the leaders and followers, which depend on the existing technological distance, measured by the number of rungs on the quality ladder, between the two parties.

When the distance between a leader and a follower is one rung apart on the quality ladder, we maintain the GH assumption that both the leader and follower can innovate with the same technology. That is, if a firm hires an amount λ of skilled workers to perform R&D, the firm experiences an arrival of a successful innovation at a Poisson rate $\Gamma(\lambda)$ given by

$$\Gamma(\lambda) = \chi\lambda, \text{ for } \chi > 0,$$

where χ is a parameter that governs the productivity in the R&D sector. The innovation technology displays constant returns at the firm's level.

When the technological distance between a leader and a follower is two rungs or steps apart, the follower can no longer innovate with the same technology as that used by the leader. Assume that the cost of innovation to the follower who is two steps behind the leader is high enough that the follower stops innovating completely. In Appendix B, we show that this assumption is without loss of generality, because, under the assumption of linear innovation technologies, either a step-by-step catch-up or a fast catch-up process as in Acemoglu and Akcigit (2012) will give us the result that followers who are two or more steps behind the leader optimally choose not to innovate.

This structure of the innovation technology is meant to capture the idea that leapfrogging becomes increasingly difficult for followers when their technological distance to the industry leader increases. There are two complementary interpretations for this assumption.

The first is that every state-of-the-art version of a product incorporates elements from the previous versions, which are patented. If the follower's technology, on which he owns patents, is not far from that of the leader (i.e. when the follower is only one step behind), then the follower is able to invent the new quality for the product without having to incorporate in this new quality any element for which only the leader owns a patent. However, as the Apple-Samsung case indicates, patents can impose substantial legal and uncertainty costs for challengers. In fact, when the followers own patents on very obsolete technologies (i.e. when the follower is two steps behind the leader), then it is not possible for the follower to invent the state-of-the-art quality without having to incorporate elements that have already been patented by the leader. But, as in GH, leaders do not have any incentive to grant a license to a follower.

The second is that knowledge spillover does take place, but it takes time. If a leader is a lot more advanced in his stock of knowledge, then it takes a longer time for the knowledge spillover to complete. In this case, followers can fall farther behind the leader in the amount of R&D knowledge they can muster when the distance to the leader is larger.

When leaders and followers are one step apart, there are potential incentives for both to innovate. Followers innovate to replace the leader, as in GH. Leaders may also want to innovate for the pure goal of distancing themselves further from the followers. As the distance grows, innovation costs for followers rise and the followers stop threatening. The incumbent's leadership will then be secured for a long period of time through innovation in the current period. We refer to this strategy of the leader as an *endpoint* strategy.

We assume that when a leader is two steps ahead of a follower, the distance is reduced to one step at an exogenous (small) rate $\tau > 0$. When rising R&D costs are interpreted as driven by legal constraints imposed by patents, then τ can be thought as a policy variable that controls the legal term of patents. When rising R&D costs are tied to lack of full knowledge spillover, then τ indicates the frequency at which the spillover occurs. It is worth-noting that when we let τ go to infinity, we are back to the GH world where there is instant spillover of the innovation technology and leaders and followers can be at most one step apart.

We say that an industry is in the *contestable* state if the distance between a leader and followers in that industry is equal to one step, and in the *non-contestable* state if the distance between a leader and followers in that industry is two steps. We indicate with $\alpha(t) \in [0, 1]$ the share of industries that at a given time t are in the contestable state. Since innovation only takes place in a contestable state, we call $\alpha(t)$ the *extensive margin* of innovation in the economy.

Mathematically, the combination of two types of firms (leader l or follower f) and of two possible distances (1 or 2) between firms, gives rise to four value function $V^j_\Delta(t)$, for $j \in \{l, f\}$ and $\Delta \in \{1, 2\}$. When it does not create confusion, we omit the indication of the dependence of variables on time. Our four value functions, at any point of differentiability, satisfy

$$rV^l_2 = \Pi + \tau(V^l_1 - V^l_2) + \dot{V}^l_2 \tag{8}$$
$$rV^f_2 = \tau(V^f_1 - V^f_2) + \dot{V}^f_2 \tag{9}$$
$$rV^l_1 = \max_{\lambda^l \geq 0} \Pi - \omega\lambda^l + \chi\lambda^l(V^l_2 - V^l_1) + \chi\lambda^f(V^f_1 - V^l_1) + \dot{V}^l_1 \tag{10}$$
$$rV^f_1 = \max_{\lambda^f \geq 0} -\omega\lambda^f + \chi\lambda^f(V^l_1 - V^f_1) + \chi\lambda^l(V^f_2 - V^f_1) + \dot{V}^f_1. \tag{11}$$

The free entry of the followers implies that

$$V^f_1 = V^f_2 = 0, \forall t.$$

2.3 Characterization of the Steady States

In equilibrium, R&D strategies are symmetric across all industries. We focus on Markov equilibria. Therefore, at any given point in time, efforts $\lambda^l(t)$ and $\lambda^f(t)$ by leaders and followers, and the corresponding intensive margin $\Lambda(t)$, are the same across all industries in the contestable state. The evolution of the extensive margin is

$$\dot{\alpha}(t) = (1 - \alpha(t))\tau - \alpha(t)\chi\lambda^l(t). \tag{12}$$

The aggregate number of rungs on the quality ladder achieved at time t, $S(t) = \int_{[0,1]} s_t(i)di$, evolves according to

$$\dot{S}(t) = \chi H(t) \equiv \chi \Lambda(t)\alpha(t), \tag{13}$$

where $H(t)$ is defined to be the total amount of skilled R&D labor employed at time t.

The definition of equilibrium in this model is standard.

Definition 1. *An equilibrium is given by prices* $\{r(t), w(t), p_t(i), \omega(t)\}_{t=0}^{\infty}$, *innovation rates by leaders and followers* $\left\{\lambda^{l*}(t), \lambda^{f*}(t)\right\}_{t=0}^{\infty}$, *functions* $\{E^*(t),\ L^*(t),\ \Lambda^*(t),\ S^*(t),\ Y^*(t),$ $\Pi(t)^*,\ \alpha^*(t)\}_{t=0}^{\infty}$ *for aggregate expenditure, supply of production labor, supply of R&D labor, aggregate quality, output, profits and the extensive margin of innovation, such that*

i) Given prices and the evolution of $\Pi(t)^*$, *the innovation rates* $\lambda^{l*}(t)$ *and* $\lambda^{f*}(t)$ *solve firms' innovation game.*

ii) Given aggregate expenditure $E^*(t)$ *and normalized wages* $w(t) = 1$, Y_t^* *and* $p^*(t) = m$ *are the optimal output and price level chosen by leaders in any industry. Correspondingly* $\Pi^*(t) = (m-1)Y^*(t)$ *are the profits of leaders.*

iii) Given prices and the evolution of aggregate profits, then $E(t)$, $L^*(t)$ *and* $\Lambda^*(t)$ *are, respectively, the optimal expenditure, and the optimal production and R&D labor supplies of households.*

iv) Given $\Lambda^*(t)$, *the wage premium* $\omega(t)$ *of firms in the contestable state satisfies (A-1). Given* $\lambda^{l*}(t)$ *and an initial condition* $\alpha(0)$, *the extensive margin* $\alpha^*(t)$ *satisfies (12). Given an initial condition* $S(0)$ *and the evolution of* $H^*(t) = \alpha^*(t)\Lambda^*(t)$, *the aggregate quality* $S^*(t)$ *satisfies (13).*

v) Markets clear, i.e. $L^*(t) = Y^*(t)$, $E^*(t)/m = Y^*(t)$, $\Lambda^*(t) = \lambda^{f*}(t) + \lambda^{l*}(t)$, $H^*(t) = 1 - L^*(t)$.

Note that, since in equilibrium the quantities of all goods are the same and equal to the production labor input, $d_t(i) = L(t)$, the log aggregate consumption can be written as

$$\log C(t) = \int_{[0,1]} \log(q_t(i)d_t(i))di = \int_{[0,1]} \log \gamma^{s_t(i)}di + \log L(t) = \log(\gamma)S(t) + \log L(t).$$

The growth rate of consumption is therefore

$$\frac{\dot{C}(t)}{C(t)} = \log(\gamma)\dot{S}(t) + \frac{\dot{L}(t)}{L(t)} = \log(\gamma)\chi H(t) + \frac{\dot{L}(t)}{L(t)} = \log(\gamma)\chi\alpha(t)\Lambda(t) + \frac{\dot{L}(t)}{L(t)}.$$

The growth rate of aggregate consumption is then given, in equilibrium, by the sum of the growth rate of the production labor input and the growth of the aggregate quality, times the log of the innovation size. We refer to the quantity $g(t) = \log(\gamma)\dot{S}$ as the *rate of technological growth*.

The balanced growth path of this model is an equilibrium where aggregate consumption and quality, $C(t)$ and $S(t)$, grow at the same rate g. Next we show that, depending on the parameter values, the equilibrium economy can display up to three steady states. To distinguish the three possible steady states, we label them using subscripts H, M or L, which indicate whether a steady state is characterized by a high, medium or low value for the extensive margin of innovation, α.

2.3.1 The H Steady State

The highest feasible steady state value for α^* is 1. That steady state is compatible with followers innovating only, and thus $\lambda^{l*} = 0$ and $\lambda^{f*} > 0$. As mentioned above, by taking $\bar{\Lambda}$ large enough, we can make sure that in a neighborhood of the steady state $\Lambda^*(t) = \lambda^{f*}(t) < \bar{\Lambda}$, giving the skill premium $\omega(t)$ equal to 1. Hence, the first order condition for λ^f in a neighborhood of a H steady state imply that

$$V_1^l(t) = \frac{1}{\chi}.$$

The condition above implies that, in a neighborhood of the H steady state, $\dot{V}_1^l = 0$. Since $\lambda^{l*}(t) = 0$ in the neighborhood of the steady state, a straightforward substitution in the definition of V_1^l gives

$$\frac{r(t)}{\chi} + \lambda^f(t) = \Pi(t). \tag{14}$$

Combining the above equation with the facts that $\Pi = (m-1)L$, $\lambda^f = (1-L)/\alpha$ and $r = \rho + \dot{L}/L$, we obtain

$$\frac{\dot{L}}{L} = \chi\left[\left(m - 1 + \frac{1}{\alpha}\right)L - \frac{1}{\alpha}\right] - \rho. \tag{15}$$

Equation (15) defines the evolution of the economy around the H steady state, together with the condition

$$\dot{\alpha} = \tau(1-\alpha). \tag{16}$$

The H steady state is then characterized by

$$\alpha_H^* = 1; \tag{17}$$

$$L_H^* = \frac{\rho+\chi}{\chi m}; \tag{18}$$

$$\lambda_H^{f*} = \frac{(m-1)\chi - \rho}{\chi m}. \tag{19}$$

Linearizing the system, (15) and (16), we can show that the H steady state is a saddle. The non-negativity of λ_H^{f*} requires that $m > 1 + \frac{\rho}{\chi}$.[3]

The value to a leader who is hypothetically two steps ahead in the H steady state can be computed as

$$V_2^{l*} = \frac{(m-1)\frac{\rho+\chi}{\chi} - \tau}{\chi(\rho+\tau)}$$

To ensure that indeed leaders do not want to innovate to be two steps ahead, we define $\overline{M}$ such that $V_{2,H}^{l*} = 2/\chi$ and require $m < \overline{M}$ for the existence of the H steady state. One can show that

$$\overline{M} = \frac{\rho+\chi}{\chi-\rho-\tau}. \tag{20}$$

For the above to be a meaningful condition, we assume $\chi > \rho + \tau$, which is in the spirit of exploring the implications of a small τ.

[3] For $m < 1 + \frac{\rho}{\chi}$, the steady state will be characterized by $\alpha_H^* = L_H^* = 1$ and $\lambda^{f*} = 0$, a case that we rule out for the lack of relevance.

2.3.2 The M and L Steady States

In the M and L steady states, both leaders and followers innovate in the contestable state. The first order conditions for λ^l and λ^f give $V_2^l(t) = 2\omega(t)/\chi = 2V_1^l(t)$. Substituting these conditions into the value functions and after appropriate calculations we obtain the two equations:[4]

$$\begin{cases} \Pi = (2\lambda^f - \frac{\tau}{\chi})\omega \\ \dot{\omega} = (r + \tau - \chi\lambda^f)\omega. \end{cases} \tag{21}$$

The M and L steady states differ in whether the supply of R&D labor is exhausted or not. In the M steady state, by definition, the industry-level R&D labor supply constraint is not binding. Hence, $\lambda_M^{l*} + \lambda_M^{f*} = \Lambda_M^* < \overline{\Lambda}$ and $\omega_M^* = 1$. In contrast, in the L steady state, the industry-level skilled labor supply binds at $\overline{\Lambda}$ and $\omega_L^* > 1$. In either steady state, the interest rate is ρ and the second equation of (21) implies followers innovate at the same intensity,

$$\lambda_M^{f*} = \lambda_L^{f*} = \frac{\rho + \tau}{\chi} \tag{22}$$

In the M steady state, we can solve out the production labor from the first equation in (21), together with $\Pi_M^* = (m-1)L_M^*$ and $\omega_M^* = 1$:

$$L_M^* = \frac{2\lambda_M^{f*} - \frac{\tau}{\chi}}{m-1} = \frac{2\rho + \tau}{\chi(m-1)}. \tag{23}$$

In the L steady state, since the skilled labor supply binds, we have $\lambda_L^{l*} = \overline{\Lambda} - \frac{\rho+\tau}{\chi}$. The evolution of the extensive margin, (12), then implies that in the steady state,

$$\alpha_L^* = \frac{\tau}{\tau + \chi\lambda_L^{l*}} = \frac{\tau}{\chi\overline{\Lambda} - \rho}.$$

This, in turn, pins down the production labor and profit in the steady state,

$$L_L^* = 1 - \alpha_L^*\overline{\Lambda} = \frac{(\chi - \tau)\overline{\Lambda} - \rho}{\chi\overline{\Lambda} - \rho};$$
$$\Pi_L^* = (m-1)\frac{(\chi - \tau)\overline{\Lambda} - \rho}{\chi\overline{\Lambda} - \rho}.$$

[4] For detailed derivations, see Appendix A.2.

With these inputs, we can solve out the value to a leader who is one step ahead from (10),

$$V_{1,L}^{l*} = \frac{\Pi_L^*}{\rho + \chi\lambda_L^{f*}} = (m-1)\frac{(\chi-\tau)\overline{\Lambda}-\rho}{(2\rho+\tau)(\chi\overline{\Lambda}-\rho)}.$$

For the L steady state to exist, $V_{1,L}^{l*} = \frac{\omega_L^*}{\chi} > \frac{1}{\chi}$. Define the m such that $V_{1,L}^{l*} = \frac{1}{\chi}$ as $\underline{\mathrm{M}}$. Therefore, the existence of the L steady state requires $m > \underline{\mathrm{M}}$, which is given by

$$\underline{\mathrm{M}} = 1 + \frac{(2\rho+\tau)(\chi\overline{\Lambda}-\rho)}{\chi\left((\chi-\tau)\overline{\Lambda}-\rho\right)}. \tag{24}$$

Comparing $\underline{\mathrm{M}}$ and $\overline{M}$, we have $\underline{\mathrm{M}} < \overline{M}$ if and only if

$$\overline{\Lambda} > \frac{\rho+\tau}{\chi},$$

an assumption that we maintain by letting $\overline{\Lambda}$ be sufficiently large.

2.4 Steady States: Discussion

The results derived in the previous section can be collected as follows:

Proposition 1. *There are two constants $\underline{M} < \overline{M}$, defined by (20) and (24), such that*

i) For $m < \underline{M}$, only the H steady state exists. For $m > \overline{M}$ only the L steady state exists. For $m \in [\underline{M}, \overline{M}]$ the steady states H, M, L all exist. Over the interval where they exist, α_M^* is increasing in m.

ii) The H and the L steady states have the saddle-path property, while the M steady state is a source. In particular, if $m > \overline{M}$, then, for any initial condition $\alpha(0)$, the economy always converges to the L steady state.

Proof. See Appendix A.2. □

It is worth pointing out that we have $\alpha_H^* > \alpha_M^* > \alpha_L^*$ when all three steady states exist, and hence our naming of these steady states. Moreover, we not only can rank the steady states by the extensive margin of innovation, they are also ranked by the steady state growth rate (Corollary 1).

Corollary 1. *For $m \in (\underline{M}, \overline{M})$, the steady state growth rate satisfy $g_H^* > g_M^* > g_L^*$. Moreover, the unique steady state growth g_L^* associated with $m > \overline{M}$ is smaller than any steady state growth rates corresponding to $m \in (\underline{M}, \overline{M})$.*

Proof. See Appendix A.3. □

The mechanism behind the structure of the steady states is based on the joint effect of two simple properties of the model: the effect of leaders' innovation on aggregate growth, and the effect of the net present value of monopoly on leaders' incentives to innovate.

First, greater innovation by leaders is associated with lower long run technological growth – the result outlined in Corollary 1. This is not surprising, since leaders' innovation is motivated by endpoint strategies, whose sole goal is to discourage innovation. The positive effect on growth from a higher intensive margin of innovation carried out by leaders in contestable states is more than offset by the greater fraction of industries that, in the long run, end up in non-contestable states with zero innovation. To see this, recall that the rate of technological growth is proportional to the product of the extensive and intensive margins. Using (12) to calculate the steady state value of the extensive margin, we obtain

$$g^* = \log(\gamma)\chi\alpha^*\Lambda^* = \log(\gamma)\chi\frac{\tau}{\tau + \chi\lambda^{l*}}\Lambda^* = \log(\gamma)\chi\frac{\tau}{\tau + \chi\lambda^{l*}}(\lambda^{f*} + \lambda^{l*})$$

A larger innovation rate by leader is then associated with lower aggregate growth, provided that $\chi\lambda^{f*} > \tau$. This latter condition, which in our case always holds in equilibrium, simply requires that the probability of a successful innovation by followers in the contestable state is greater than he exogenous probability of a spillover (or of a patent expiration) in the non-contestable state.

Second, a higher net present value of monopoly power increases leaders' incentives to innovate. This is intuitive, since a larger leadership value triggers more effort to secure it by means of endpoint strategies. Indeed, it is straightforward to show that the incremental value $V_2^l - V_1^l$ to a leader who successfully innovates is given by[5]

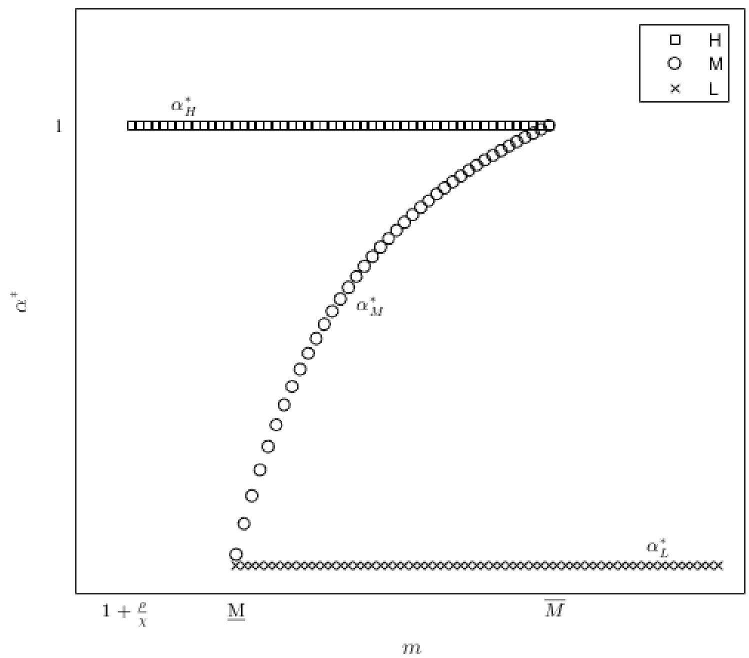
$$V_2^l - V_1^l = \frac{\chi\lambda^f}{(r + \chi\lambda^f)(r + \tau)}\Pi.$$

Leaders' endpoint strategies are incentivized when monopolist's profit Π is higher, or follow-

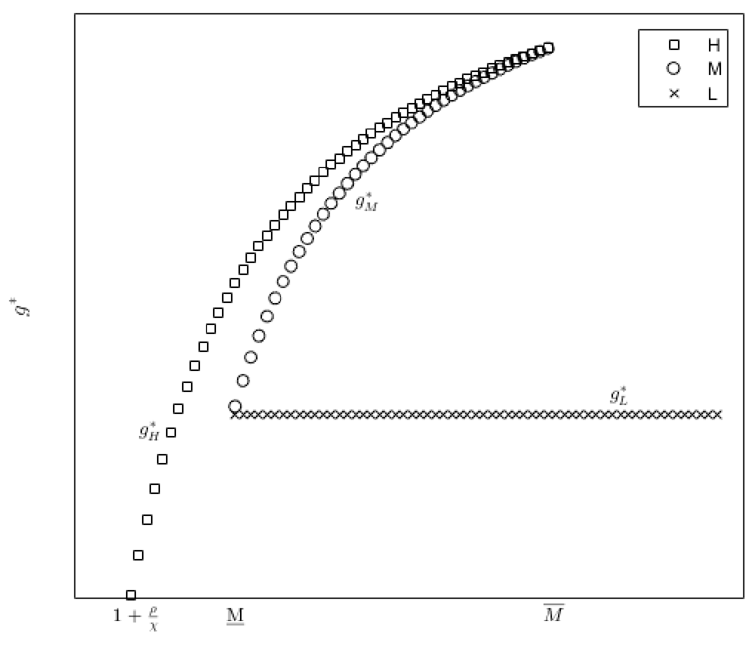
[5]At the steady state, (10) implies $V_1^l = \frac{\Pi}{r+\chi\lambda^f}$ and (8)-(10) implies $V_2^l - V_1^l = \frac{\chi\lambda^f V_1^l}{r+\tau}$. Combining the two, we obtain the expression $V_2^l - V_1^l$ in the paper.

Figure 2: Extensive Margin of Innovation and Growth Rate in the Steady States

(a) Steady State Extensive Margin of Innovation, α_i^* for $i = H, M, L$



(b) Steady State Growth Rate, g_i^* for $i = H, M, L$



Note: This figure illustrates the structure of the steady states of the model. In particular, it shows how the steady state extensive margin of innovation α^* and the steady state growth rate g^* vary as we vary the markup parameter m. For a discussion, see Section 2.4.

ers' innovation rate λ^f is higher, or the interest rate r is lower, or the externality intensity τ (frequency of patent expiration) is lower. Profits are higher, for instance, when markups m are larger, which explains the result of Proposition 1, depicted in Figure 2. Panel (a) of the figure plots the steady state extensive margin of innovation α^* against m. As discussed in the next section, for any initial $\alpha(0) < \alpha_M^*$, the equilibrium path converges to the L steady state. As the figure shows, if markups m are too large, i.e. $m > \bar{M}$, then incentives for leaders to innovate are so strong that for any initial condition the economy converges to L, which is the steady state with lowest long run growth (Panel (b)). A similar reasoning explains why when knowledge spillover (or patents expiration) are infrequent, then only the L steady state exists (notice in fact that $\bar{M}$ is increasing in τ). Instead, a higher discounting r reduces the present value of profits, and thus discourages R&D by leaders, while higher innovation by followers, by increasing the threat to the incumbents, strengthens their incentives to play end point strategies.

For values of $m \in [\underline{M}, \bar{M}]$, the model features multiple steady states. The key to understanding this result is to link, in general equilibrium, the two properties discussed above. Fix a given $m \in [\underline{M}, \bar{M}]$ and begin by assuming that the economy is in a high growth steady state. Since a large fraction H of the labor input is devoted to R&D, the production labor and period profits are low. With low profits, incentives for leaders to innovate fall. For similar reasons, innovation intensity for followers λ^{f*} is also reduced (in an H steady state the extensive margin of innovation is large, but the intensive margin is small).[6] This further depresses leaders' innovation incentives. Finally, while under the log-utility assumption considered so far steady state interest rate $r = \rho$ is independent of the steady state growth rate g, Section 4.1 shows that, when the elasticity of intertemporal substitution is greater than 1, then a larger growth rate raises the equilibrium interest rate, further dampening innovation incentives for leaders. In conclusion, if the economy is at a H steady state, then general equilibrium effects discourage innovation by leaders, and since R&D by leaders is negatively associated with long run growth, then the high growth steady state is self-confirmed. A similar line or reasoning can be followed, for instance, to self-confirm an initial position at the L steady state. In particular, in the region with multiple steady states, the L equilibrium represents a "monopolistic growth trap." An economy that starts from a non-contestable position $\alpha(0) < \alpha_M^*$ will suffer general equilibrium effects that keeps it there for ever.

[6] If we compare (22) with (19) we note in fact that $\lambda_H^{f*} < \lambda_M^{f*} = \lambda_L^{f*}$ for $m < \bar{M}$.

2.5 The Equilibrium Saddle Paths

So far our discussion has focused on the steady states of the model. In this section, we simulate the entire equilibrium saddle path picked by the initial condition of the extensive margin of innovation, $\alpha(0)$, in the model parametrized by a set of empirically plausible parameters.

We can characterize the equilibrium path by the following system of differential equations. From an initial $\alpha(0) \in (\alpha_L^*, \alpha_M^*)$ that is sufficiently close to α_M^*, $(\alpha(t), L(t))$ evolves according to

$$\begin{cases} \dot{\alpha}(t) = \rho\alpha(t) + \frac{2\rho+\tau}{m-1} + \tau - \chi \\ \dot{L}(t) = \frac{2\rho+\tau}{\chi(m-1)} \end{cases} . \tag{25}$$

while $\lambda^l + \lambda^f < \overline{\Lambda}$. As soon as $\overline{\Lambda}$ starts to be binding, the system switches to

$$\begin{cases} \dot{\alpha}(t) = \tau - \alpha(t)(\chi\overline{\Lambda} - \rho) \\ \dot{L}(t) = 1 - \alpha(t)\overline{\Lambda} \end{cases} .$$

The economy converges to the L steady state.

From an initial $\alpha(0) \in (\alpha_M^*, \alpha_H^*)$ that is sufficiently close to α_M^*, both leaders and followers innovate. The system evolves according to (25) until leaders are indifferent between innovating and not innovating, i.e. until $V_2^l(t) = 2$. After that, leaders stop innovation and the equilibrium jumps to the saddle path that converges to the H steady state where only followers innovate. The system evolves according to:

$$\begin{cases} \dot{\alpha}(t) = (1 - \alpha(t))\tau \\ \dot{L}(t)/L(t) = \chi\left((m-1)L(t) - \frac{1-L(t)}{\alpha(t)}\right) - \rho \end{cases} .$$

We simulate the model under the parameters given in Table 1. A period in the model is one year. We set the subjective discount rate to 0.02. The skilled labor supply cap, $\overline{\Lambda}$, is calibrated to the percentage of college graduates among adult population in the US in 2017, not all of whom need work in the R&D sector. The rate of destruction of the non-contestable state, τ, is taken to be 0.05, which is the inverse of the term of patents (20 years). We set m to be in the interval $[\underline{M}, \overline{M}]$, so all three steady states exist. The step size of the innovation, γ, is chosen to ensure reasonable growth rates of the economy.

The saddle equilibrium path that converges to the L steady state is illustrated in Panel

(a) of Figure 3. The initial condition of the extensive margin of innovation, $\alpha(0)$, is chosen to be just below α_M^*. Initially, around the M steady state, both leaders and followers innovate. Since the extensive margin of innovation decreases, leaders must increase the intensive margin of innovation. As long as the skilled labor supply is not binding in contestable industries, the two margins that evolve in opposite directions counteract each other perfectly, so the rate of technological growth pinned down by the aggregate level of innovation, $g(t)$, is constant and so is the size of the production sector, $L(t)$. As soon as the skilled labor supply binds in contestable industries, the decline in the extensive margin takes over and the aggregate level of innovation declines and the production sector expands. Admittedly, the kink in $g(t)$ is driven by the fixed supply of skilled labor necessitated by the linearity of the model. In Section 4.2, where we relax the assumption of the linear R&D cost structure, $g(t)$ will evolve smoothly to the L steady state.

Panel (b) of the same figure shows the saddle equilibrium path that converges to the H steady state. Starting from an initial extensive margin of innovation just above the M steady state, the extensive margin of innovation increases, the intensive margin innovation decreases due to leaders innovating less, while the aggregate innovation stays constant, until the moment leaders no longer find it profit to innovate. At that point, the equilibrium jumps to the saddle path that converges to the H steady state. The extensive margin $\alpha(t)$, moves continuously, though its rate of change has a kink at that point, while the production labor, $L(t)$, has a discontinuous jump. Afterwards, the extensive margin keeps increasing until all industries are contestable and we arrive at the H steady state.

It is conceivable that for an initial condition $\alpha(0)$ that is close to α_M^*, changing markup policies or patent policies which affect m and τ, can have far-reaching long-run implications if the policy change alters the steady state that the economy is heading to.

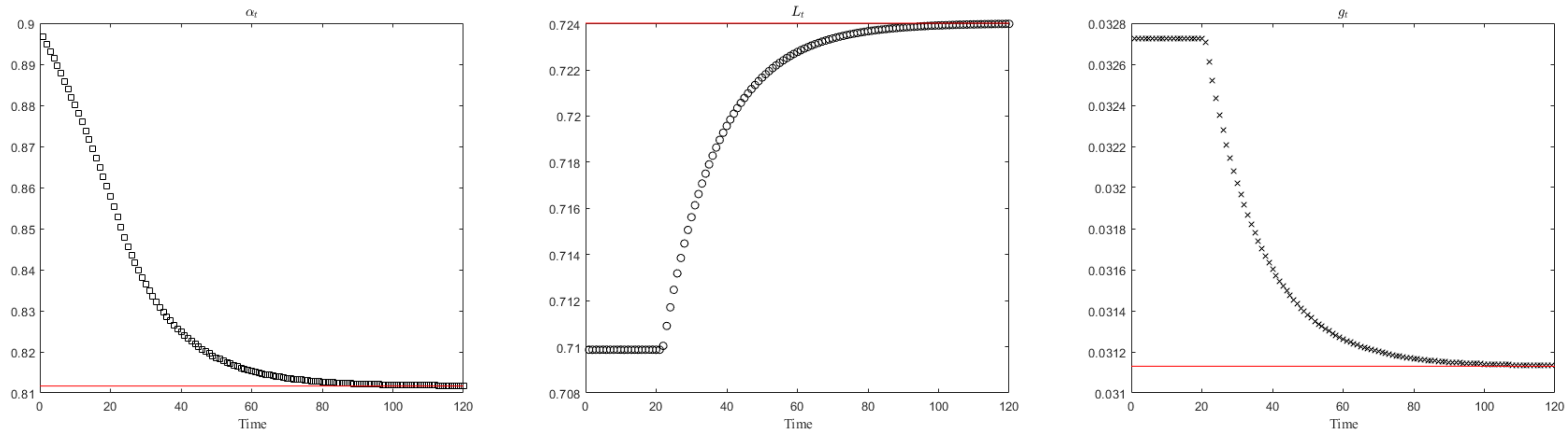
Table 1: Parameter Values

Parameter	Value	Justification
ρ	0.02	Convention
$\overline{\Lambda}$	0.34	Pct. of college graduates among adult population
τ	0.05	Term of patents, 20 years
χ	0.24	Ensure the existence of path to the H
m	1.53	The average of $\overline{M}$ and $\underline{M}$
γ	1.60	Ensure $\gamma > m$ and reasonable consumption growth rate

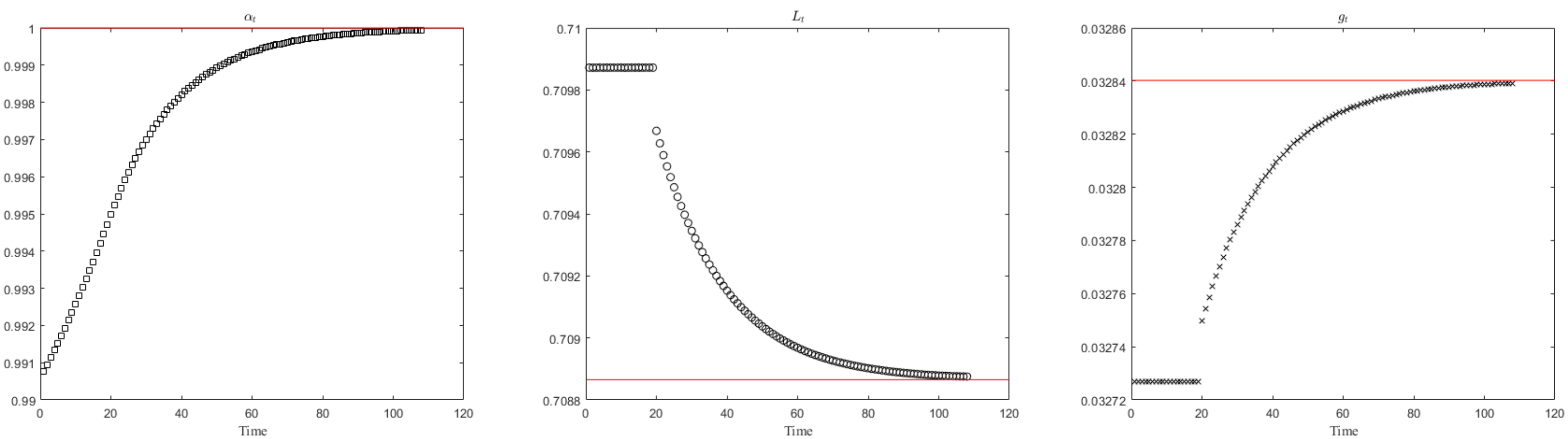
Note: This table reports the parameter values we use in the simulation of the baseline model and their justifications. For a discussion, see Section 2.5.

Figure 3: Equilibrium Saddle Paths

(a) Saddle Path to the L Steady State



(b) Saddle Path to the H Steady State



Note: This figure shows the simulated equilibrium paths to the L and the H steady state. In particular, we plot the evolution of the extensive margin of innovation α_t, the amount of production labor L_t, and the steady state growth rate g_t on their saddle paths to the steady state. The model is simulated under the parametrization given by Table 1. For a discussion, see Section 2.5.

3 Policy Implications

Should Medicare be allowed to bargain for better deals with drugs providers, effectively reducing the markup for pharmaceutical companies? What is the effect of longer patents' duration? These are all common policy questions that, in our model, involve setting the parameters m and τ. As shown in the previous section, in general the growth rate of the economy responds in a non-linear way to changes in these parameters. Moreover, the effects in the long run can even depend on the initial condition of the economy. The legal framework thus provides a powerful set of constraints on the ability of a country to innovate and develop (Parente and Prescott, 2002).

This section explores the dynamic evolution of the economy under two policy experiments, one involving a change in m and one a change in τ. To make our experiments more striking, We look at knife-edge cases where the economy's initial condition $\alpha(0)$ is around an M steady state.

3.1 Raising the Markup Ceiling

In the first policy experiment, we raise the markup m slightly from 1.5237 to 1.5248 and simulate the equilibrium path from the same initial condition $\alpha(0) = 0.9162$ under the two different policy environments. The parametrization is otherwise identical to that in Table 1. The simulated equilibrium paths are found in Figure 4. Panels (a) to (c) illustrate the equilibrium behavior of the extensive margin of innovation $\alpha(t)$, the aggregate production labor $L(t)$, the rate of technological growth $g(t)$. Panel (d) in the same figure shows the ratio of the aggregate consumption in the low-m environment to the aggregate consumption in the high-m environment. To ease the reading of the figure, we use solid red to describe the low-m economy and hollow black to describe the high-m economy.

In the low-m economy, the initial condition $\alpha(0)$ is above the M steady state level of extensive margin, α_M^{*pre}, setting the economy on the saddle path to the H steady state. The H steady features innovation only by followers and in all industries which add up to a high level of aggregate innovation and growth. However, lowering m, ever so slightly, at the initial state of the economy increases the M steady state level of the extensive margin, α_M^{*post}, which completely changes the equilibrium path of the economy. In the high-m environment, the economy converges to the L steady state, featuring a much lower fraction of industries being contestable and a lower level of aggregate innovation and growth.

To evaluate the aggregate consequence of such a policy change, we track the aggregate consumption in the high-m environment relative to that in the low-m environment period by period in Panel (d). It is noteworthy that, even in our model, raising markups for leaders produces higher aggregate consumption in the short run, before the negative long-run effects kick in. The reason for the diverging short-run and long-run welfare implications of raising the markup ceiling is as follows. As monopoly profits are increased under a larger m, leaders respond to it by raising their effort in innovation while followers' optimal intensive margin of innovation remain unchanged.[7] On the other hand, higher intensive margin of innovation by the leaders increases the speed at which an industry escapes the contestable state and reduces the extensive margin of innovation (Panel (a)). Therefore, the intensive and extensive margin of innovation move in opposite direction in the short run after the policy change. Panels (b) and (c) tell us that the first of the two effects dominates in the short run so that aggregate innovation increases and aggregate labor employed in production decreases. Over time, the skilled labor employed in the R&D sector in those fewer and fewer contestable industries is exhausted, so the intensive margin of innovation at the industry level cannot increase while the extensive margin keeps decreasing. In the long run, the second effect clearly dominates, canceling out any short-run gain and leading to a permanently lower consumption growth.

3.2 Lengthening Patents' Duration

We now consider another policy experiment, where the destruction rate of the contestable state, τ is reduced marginally from 0.051 to 0.05. Consider an economy with an initial condition that $\alpha(0) = 0.8576$. We plot the simulated saddle path to their respective steady state under the low-τ and the high-τ environment in Figure 5. The rest of the model parameters remain unchanged.

Before the policy that reduces τ, the initial extensive margin is above its M steady state level, α_M^{*pre}, which means with the proper $L(0)$ the economy is on the path to the H steady state. A reduction in τ increases the M steady state level of the extensive margin of innovation dramatically, which causes the initial $\alpha(0)$ to fall below it. As a consequence, the economy is set on a saddle path to the L steady state. In fact, as the system adapts to the new saddle path immediately after the policy change, the supply of skilled labor is binding from the very beginning on the high-τ equilibrium path. As a result, the extensive

[7] Recall that around the M steady state, the fixed skill supply is not binding and hence $\omega(t) = 1$. From the second equation of (21), it is implied that followers around the M steady state innovate at a constant intensity, $\lambda^f(t) = \frac{\rho+\tau}{\chi}$.

margin of innovation declines and so do the aggregate innovation and growth rate along the equilibrium path to the L steady state (Panels (a) to (c)).

Panel (d) of Figure 5 compares the aggregate consumption on the two equilibrium paths. The initial dip of the ratio below one is caused (as is in the previous experiment) by the initial contraction in $L(0)$ to reach a new saddle path upon the change in policy. In response to the increase in the value of being a monopolist, both leaders and followers increase their innovation effort, resulting in a reduction in labor used in production and hence a reduction in output or consumption initially. Since the industry-level skilled labor constraint binds from day 1, firm's increased research effort precipitates the decline in the extensive margin of innovation. Therefore, after a temporary increase of consumption relative to the high-τ economy, the force of declining extensive margin dominates and consequently consumption declines permanently relative to the high-τ economy.

4 Model Robustness

In this section, we consider two extensions of our stylized baseline model. In Section 4.1, we consider more general utility functions that belong to the constant intertemporal elasticity of substitution class. We discuss how the intertemporal elasticity of substitution affects our results. In Section 4.2, we relax the assumption of the linear cost of innovation and replace it with a quadratic cost of innovation. Furthermore, due to the convex cost assumption, we no longer need to impose a fixed supply of skilled labor any more. We show that the structure of the steady states in the model with the quadratic cost resembles that in the baseline model, therefore making sure the linearity of the baseline model does not drive our key results in any way.

4.1 Relaxing Log Utility

In the baseline model, we assume households have a log period utility function, which amounts to assuming unit intertemporal elasticity of substitution. In this section, we relax this assumption by adopting a more general class of utility functions for households, one that features the constant intertemporal elasticity of substitution:

$$\int_0^\infty e^{-\rho t} \frac{C(t)^{1-\sigma}}{1-\sigma} dt,$$

where $\frac{1}{\sigma}$ is the intertemporal elasticity of substitution. All other elements of the model remain the same as in the baseline model. The consumption Euler equation becomes:[8]

$$r(t) = \rho + \sigma \frac{\dot{E}(t)}{E(t)} + (\sigma - 1) \log(\gamma) \dot{S}(t). \tag{26}$$

In a steady state, we have a relationship between the interest rate and the rate of technological growth:

$$r^* = \rho + (\sigma - 1) g^*.$$

When we have unit elasticity, $\sigma = 1$, steady state interest rate is equal to the rate of time preference ρ, as in the baseline model.

Under the more general utility function, the steady state interest rate depends positively (negatively) on growth when σ is larger (smaller) than unity. This implies that when $\sigma > 1$, in a steady state with higher technological growth and innovation, the interest rate will also be higher. As discussed in Section 2.4, this is the second source of the multiplicity of steady states that dynamically, higher equilibrium growth rate raises the interest rate with which leaders discount future profits, and therefore reducing leaders' incentive to innovation and raising the share of contestable industries, consistent with high aggregate growth. On the other hand, when $\sigma < 1$, higher aggregate growth and innovation coexist with a lower interest rate, which then encourages leaders to innovate, decreasing the share of the contestable industries, making it more likely that the economy features a single steady state, the L steady state. In other words, the multiplicity of steady states does depend on the value of σ, however the multiplicity itself is not a particularity of the log preference in the baseline model.

We can characterize analytically the structure of the steady states for a range of σ (see Appendix C). Fix an $m \in \left(\underline{M}, \overline{M}\right)$ and all other parameters as in baseline model. There exists a range of σ, $(\underline{\sigma}_H, \overline{\sigma}_L)$ and $\underline{\sigma}_H < 1 < \overline{\sigma}_L$, in which there are three steady states, H, M, and L. As in the baseline model, the H and L steady states are saddle and the M steady state is a source. The steady state growth rate is ordered in the three steady states as their name suggest. At $\underline{\sigma}_H$, the H and M steady states coincide where leaders become indifferent between innovating and not innovating. At $\overline{\sigma}_L$, the M and L steady states coincide where the fixed skilled labor supply becomes just binding. Figure 6 illustrates the three steady states, their extensive margins of innovation and growth rates, as we vary σ. This figure

[8]The mathematical derivations are found in Appendix C.

is based on simulations of the model, keeping all parameters as in Table 1 and varying σ around unity. The red line denotes the steady states corresponding to a model with $\sigma = 1$ (i.e. the baseline model).

From the figure, for low values of σ, the only steady state that exists is the L steady state, where leaders dominate innovation and few industries are contestable. For moderate values of σ around one, multiple steady states arise. In particular, the multiplicity persists for a wider range of σ that is bigger than one, as we explain above. However, when σ becomes too big, only H steady state survives. Beyond that point, the steady state interest rate is too high to warrant innovation by leaders.

4.2 Relaxing Linear Cost of R&D

Another potential concern is whether the multiplicity of steady states, from which we derive subtle policy implications, could be driven by the linearity of the model. To address this concern, we modify the model to introduce a quadratic cost of innovation to both leaders and followers. Suppose the cost of innovation is the following:

$$C_i(\lambda) = \chi_i \lambda + \frac{1}{2}\theta_i \lambda^2, \quad i = 1, 2,$$

where $i = 1$ is for leaders and $i = 2$ is for followers. Since this effectively imposes decreasing return on innovation at the firm level, we then abandon the assumption of a fixed supply of skilled labor, which acts as a force of decreasing return on innovation at the industry level. We solve and simulate this modified model and examine if the key properties of the steady states survive these modifications.[9]

In Figure 7, we plot the steady state values of the extensive margin of innovation, α^*, and the growth rate, g^*, against different values of the markup m from the modified model. Comparing this figure to Figure 2 from the baseline model, we confirm that the structure of the steady states under the quadratic cost of innovation remains similar to that in the linear model.

[9] The mathematical derivations of the steady states in the model with quadratic costs are found in Appendix D.

5 Conclusion

Traditional "new growth theory" models of endogenous growth deliver the result that higher monopoly power - higher markup or longer patent protection - leads to higher aggregate growth. In this paper, we show that this conclusion rests crucially on the assumption of complete and instantaneous knowledge spillover to followers. We believe this assumption to be dubious, and thus we study the case where the cost for followers to leap-frog the industry's leader increases in the leader's technological advantage.

We find that under our more general setting, the equilibrium properties of the economy change dramatically. First of all, instead of being characterized by just one steady state where only followers innovate (this is the High growth steady state of the traditional Schumpeterian models), the economy now features two additional potential steady states where also leaders innovate, one with Medium and one with Low growth. The High and the Low growth steady states are both saddle path stable. The Low growth steady state is characterized by high but infrequent innovation effort by industry leaders, whose "endpoint strategy" is to acquire new patents in order to distance themselves from the followers, thus increasing the followers' innovation costs and pushing them out of the innovation race. Second, we find that when leaders are granted large monopolistic rents or long-lasting patent protection, then the economy features once again a unique steady state, but it's the Low growth steady state instead of the traditional High growth one. Allowing leaders to take advantage of excessively high markups and long patent protection is harmful to growth, as these conditions provide leaders with incentives to enact strategies aimed at stifling firms entry into their industry.

Our theoretical findings indicate that standard results of the "new growth theory" literature are not robust to the relaxation of the unrealistic assumption of complete and instantaneous knowledge spillover. Our results also provide a potential framework to interpret the recent empirical trends of increasing markups, reduced investment, and lower business dynamism.

References

Acemoglu, D. and Akcigit, U. (2012). Intellectual property rights policy, competition and innovation. *Journal of European Economic Association*, 10(1):1–42.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *Quarterly Journal of Economics*, 120(2):701–728.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60(2):323–351.

Boldrin, M. and Levine, D. K. (2008). *Against Intellectual Monopoly*. Cambridge University Press.

Budd, C., Harris, C., and Vickers, J. (1993). A model of the evolution of duopoly: Does the asymmetry between firms tend to increase or decrease? *Review of Economic Studies*, 60(3):543–573.

Grossman, G. M. and Helpman, E. (1991). Quality ladders in the theory of growth. *Review of Economic Studies*, 58(1):43–61.

Harris, C. and Vickers, J. (1987). Racing with uncertainty. *Review of Economic Studies*, 54(1):1–21.

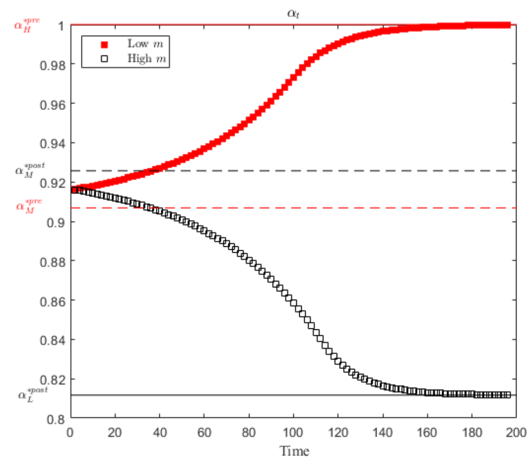
Hörner, J. (2004). A perpetual race to stay ahead. *Review of Economic Studies*, 71:1065–1088.

Klette, T. J. and Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of Political Economy*, 112(5):986–1018.

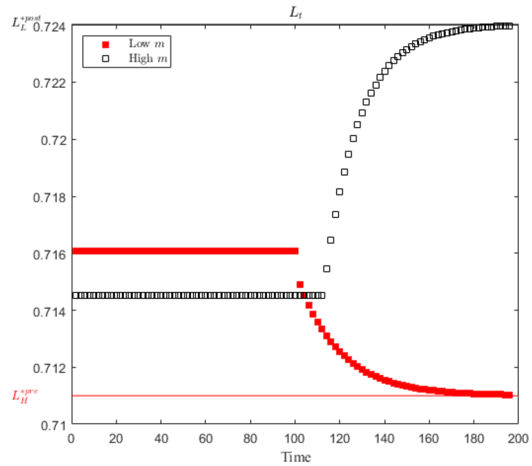
Parente, S. L. and Prescott, E. C. (2002). *Barriers to Riches*. MIT Press.

Figure 4: Policy Experiment: Raising m

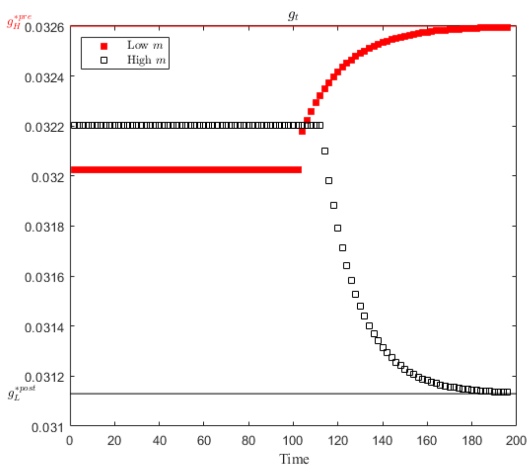
(a) Extensive margin of innovation



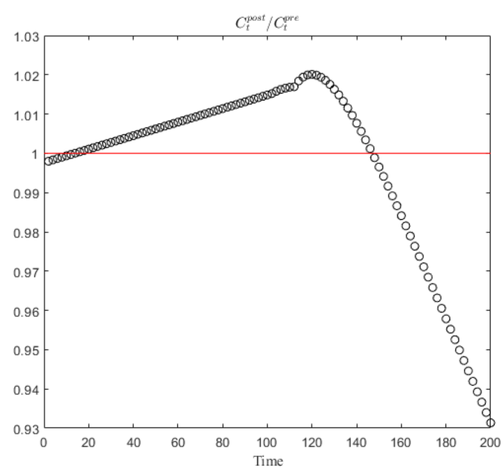
(b) Aggregate production labor



(c) Rate of technological growth



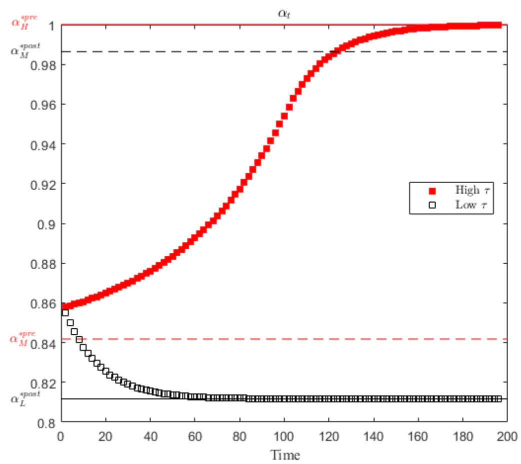
(d) Ratio of aggregate consumption



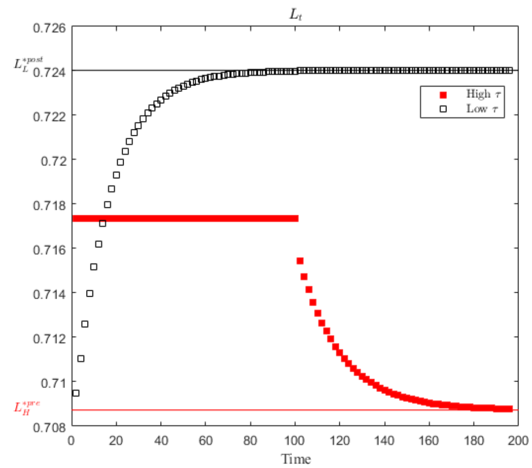
Note: This figure illustrates how the saddle path of the economy can change upon a change in the policy variable m. The red lines in Panel (a)-(c) depict the evolution of the extensive margin of innovation α_t, the production labor L_t and the growth rate g_t on the saddle path to a H steady state. Upon an increase in the markup m, the economy however lands on a saddle path converging to the L steady state, as shown by the black lines in those panels. Panel (d) shows the ratio of consumption, period by period, before to after the increase in m. For a discussion, see Section 3.1.

Figure 5: Policy Experiment: Decreasing τ

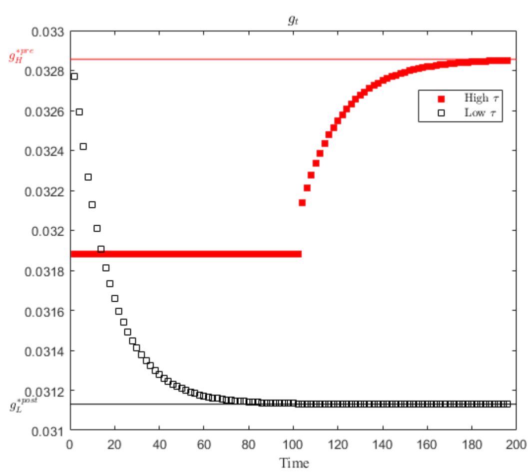
(a) Extensive margin of innovation



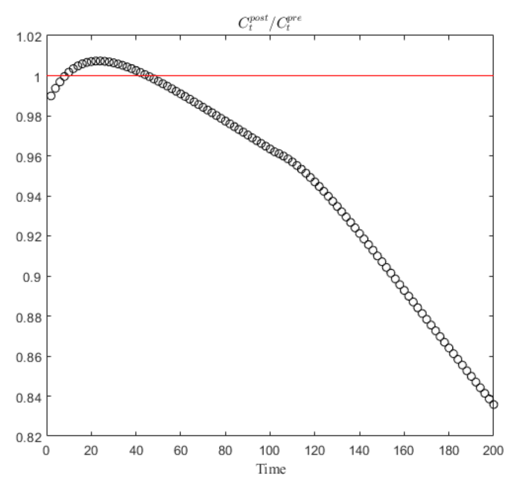
(b) Aggregate production labor



(c) Rate of technological growth



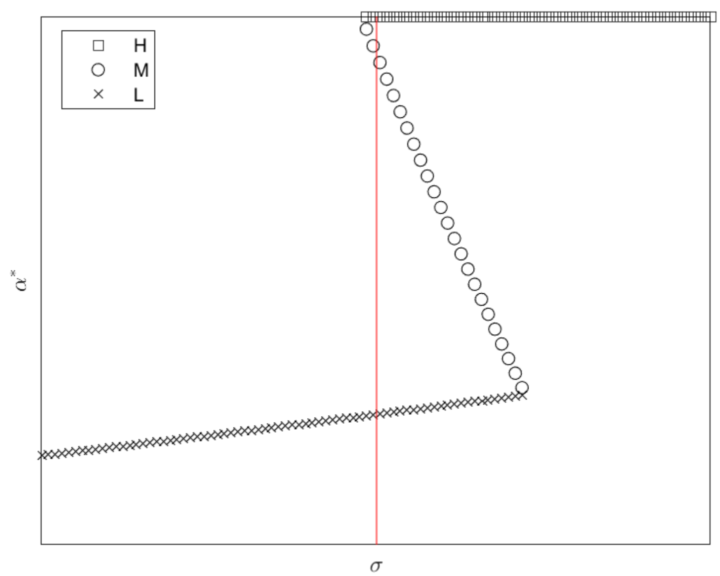
(d) Ratio of aggregate consumption



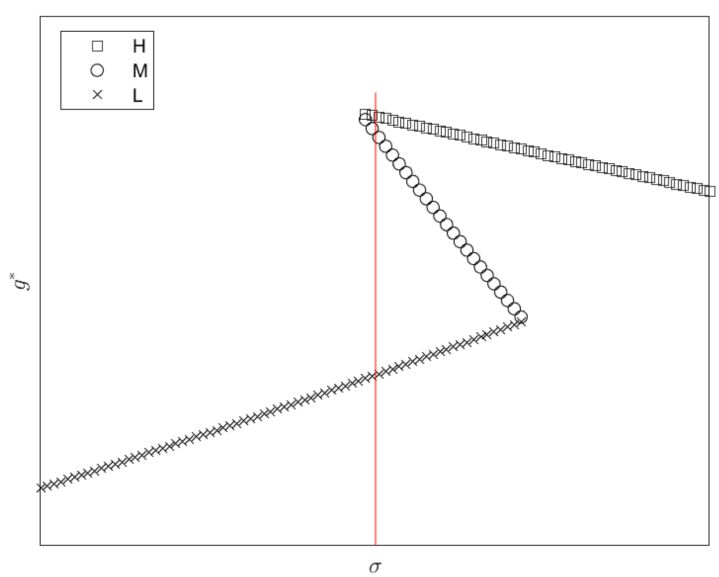
Note: This figure illustrates how the saddle path of the economy can change upon a change in the policy variable τ. The red lines in Panel (a)-(c) depict the evolution of the extensive margin of innovation α_t, the production labor L_t and the growth rate g_t on the saddle path to a H steady state. Upon a reduction in the destruction rate τ, the economy however lands on a saddle path converging to the L steady state, as shown by the black lines in those panels. Panel (d) shows the ratio of consumption, period by period, before to after the increase in m. For a discussion, see Section 3.2.

Figure 6: The Model with the Constant Intertemporal Elasticity of Substitution Preference

(a) Steady State Extensive Margin of Innovation



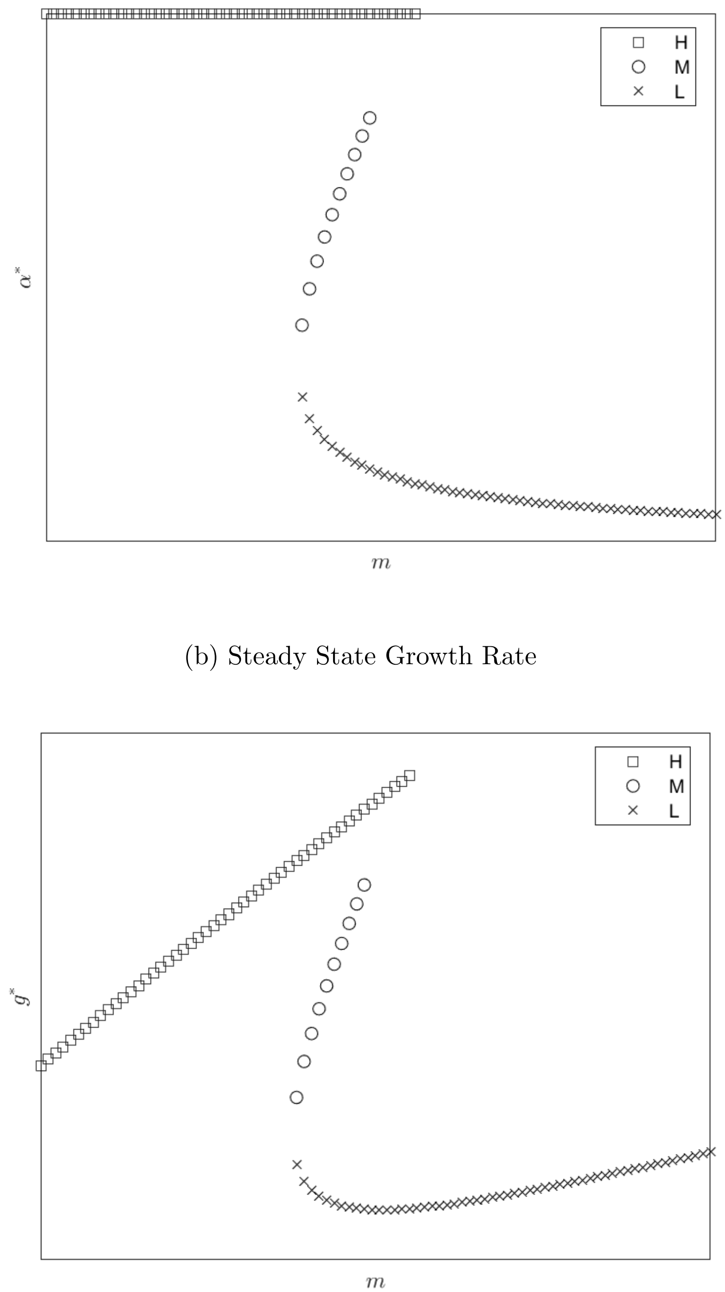
(b) Steady State Growth Rate



Note: This figure illustrates the structure of the steady states of the model extended to have constant intertemporal elasticity of substitution preference. In particular, it shows how the steady state extensive margin of innovation α^* and the steady state growth rate g^* vary as we vary the elasticity of substitution parameter σ. For a discussion, see Section 4.1.

Figure 7: The Model with Quadratic Costs of Innovation

(a) Steady State Extensive Margin of Innovation



Note: This figure illustrates the structure of the steady states of the model extended to have quadractic cost of innovation. In particular, it shows how the steady state extensive margin of innovation α^* and the steady state growth rate g^* vary as we vary the markup parameter m. For a discussion, see Section 4.2.

Appendix

A Technical Details of the Baseline Model

A.1 The Supply Function of Skilled Labor

To obtain an equilibrium when the supply function for specialized labor is perfectly elastic up to $\bar{\Lambda}$, and perfectly inelastic afterwards, we proceed in two steps. First, we postulate the existence of an exogenous supply function for the specialized labor given by

$$\omega = 1 + \theta\psi(\Lambda) \tag{A-1}$$

where $\theta > 0$ and $\psi(\cdot)$ is a C^1 function such that

$$\begin{array}{ll} \psi(\Lambda) = 0, & \text{for } \Lambda \leq \bar{\Lambda} \\ \psi', \psi'' > 0, & \text{for } \Lambda > \bar{\Lambda} \\ \psi(\Lambda) \to +\infty, & \text{as } \Lambda \to +\infty. \end{array} \tag{A-2}$$

Second, we take the limit of the resulting equilibrium as $\theta \to +\infty$. For the purpose of this paper, we take $\bar{\Lambda}$ to be an arbitrarily large constant.

A.2 The Proof of Proposition 1

The H steady state and its existence are established in the main text of the paper. Here we focus on the M and L steady states.

From the first-order conditions of the innovating leaders and followers, we have $2\omega(t)/\chi = 2V_1^l(t) = V_2^l(t)$. Therefore, $2\dot{V}_1^l = \dot{V}_2^l$. Substituting these equations into the definitions of V_1^l and V_2^l, cancelling out $\dot{V}_1^l$ and $\dot{V}_2^l$, we have

$$\Pi(t) = (2\lambda^f(t) - \tau/\chi)\omega(t), \tag{A-3}$$

which is the first equation of (21) in the paper. Cancelling out Π, we have

$$\dot{\omega}(t) = (r(t) + \tau - \chi\lambda^f(t))\omega(t),$$

which is the second equation of (21) in the paper.

From the first equation of (21), we have

$$\begin{aligned} &\frac{\dot{\Pi}}{\Pi} = \frac{\dot{L}}{L} = \frac{\dot{Y}}{Y} = \frac{\dot{E}}{E} = r - \rho = \frac{2\chi\dot{\lambda}^f}{2\chi\lambda^f - \tau} + \frac{\dot{\omega}}{\omega} = \frac{2\chi\dot{\lambda}^f}{2\chi\lambda^f - \tau} + r + \tau - \chi\lambda^f \\ \Rightarrow \quad & 2\dot{\lambda}^f = (2\lambda^f - \tau/\chi)\left(\lambda^f - \frac{\rho + \tau}{\chi}\right). \end{aligned} \tag{A-4}$$

From the expression for Π, $2\lambda^f - \tau/\chi > 0$. Then in the steady state, $\lambda^{f*} = \frac{\rho+\tau}{\chi}$. Since $\Pi^* = (m-1)(1-\alpha^*\Lambda^*)$,

$$(m-1)(1-\alpha^*\Lambda^*) = \frac{2\rho+\tau}{\chi}\omega(\Lambda^*)$$
$$\Rightarrow \quad \alpha^* = \frac{1}{\Lambda^*}\left(1 - \frac{2\rho+\tau}{\chi(m-1)}\omega(\Lambda^*)\right) \equiv \nu_1(\Lambda^*).$$

From $\dot{\alpha} = (1-\alpha)\tau - \alpha\chi\lambda^l$, we have

$$0 = (1-\alpha^*)\tau - \alpha^*\chi\lambda^{l*} = (1-\alpha^*)\tau - \alpha^*\chi(\Lambda^* - \lambda^{f*}) = (1-\alpha^*)\tau - \alpha^*\chi(\Lambda^* - \frac{\rho+\tau}{\chi})$$
$$\Rightarrow \quad \alpha^* = \frac{\tau}{\Lambda^*\chi - \rho} \equiv \nu_2(\Lambda^*).$$

The system of equations

$$\begin{cases} \alpha^* = \frac{1}{\Lambda^*}\left(1 - \frac{2\rho+\tau}{\chi(m-1)}\omega(\Lambda^*)\right) \equiv \nu_1(\Lambda^*) \\ \alpha^* = \frac{\tau}{\Lambda^*\chi-\rho} \equiv \nu_2(\Lambda^*) \end{cases}, \tag{A-5}$$

when having two meaningful solutions, define the M and L steady states.

In the limit economy, let $\theta \to +\infty$. Then,

$$\nu_1(\Lambda^*) = \begin{cases} \frac{1}{\Lambda^*}\left(1 - \frac{2\rho+\tau}{\chi(m-1)}\right) & \text{if } \Lambda^* < \overline{\Lambda} \\ -\infty & \text{if otherwise} \end{cases}.$$

In Figure A-1, we plot ν_1 (for both a finite θ and for the limit when $\theta \to \infty$) and ν_2. One can show that for finite θ, ν_1 and ν_2 have at most two crossings, because $\frac{\nu_1'}{\nu_2'}$ increases in Λ.[10]. As $\theta \to \infty$, the lower crossing occurring at the binding skilled labor constraint, defining the L steady state, $\Lambda_L^* = \overline{\Lambda}$ and the higher crossing defines the M steady state, where $\omega_M^* = 1$.

Let θ go to infinity. Varying m shifts $\nu_1(\cdot)$ up and down. Let $\underline{M}$ be the m such that $\nu_1(\cdot)$ and $\nu_2(\cdot)$ have only one intersection at $\overline{\Lambda}$. This implies that if m is even lower than $\underline{M}$, then the M and L steady states disappear. Let $\overline{M}$ be the m such that there are two intersections of $\nu_1(\cdot)$ and $\nu_2(\cdot)$, with the higher one corresponding to $\alpha^* = 1$ and the lower one corresponding to $\overline{\Lambda}$. This implies that if $m = \overline{M}$, then $\alpha_H^* = \alpha_M^* = 1$. If $m > \overline{M}$, then

[10] We have

$$\frac{\nu_1'}{\nu_2'} = \frac{\frac{2\rho+\tau}{\chi(m-1)}\theta\psi'\Lambda + \nu_1\Lambda}{\tau}\left(\chi - \frac{\rho}{\Lambda}\right)^2.$$

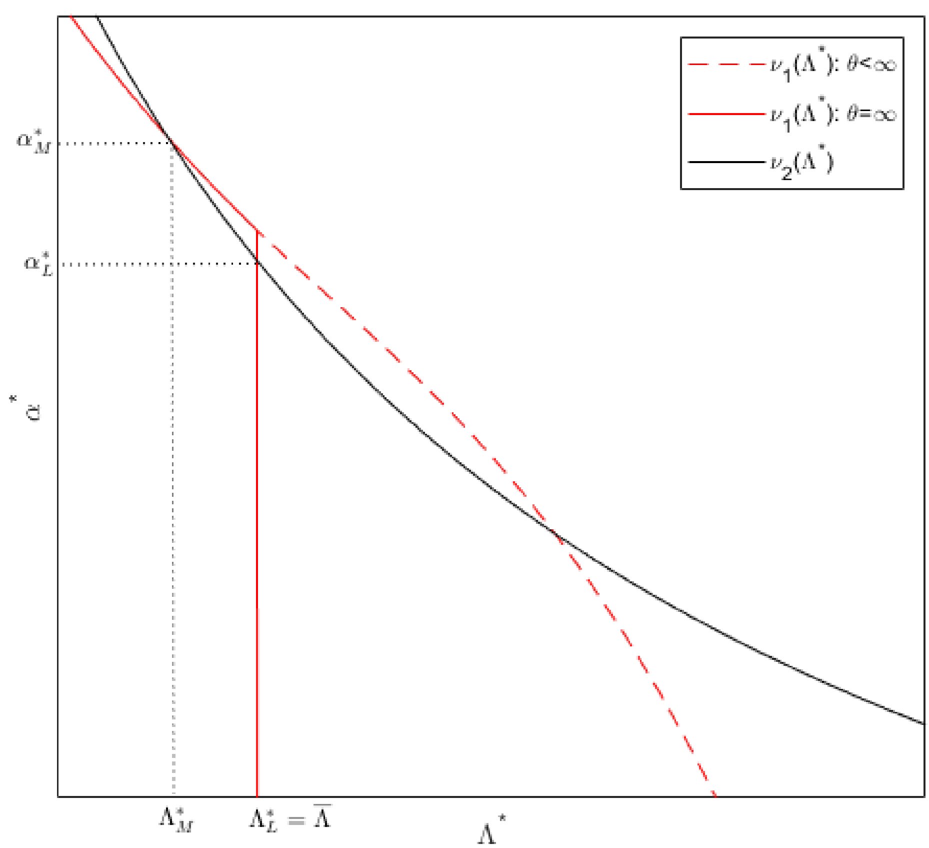
It can be shown that the first term's derivative with respect to Λ is $\frac{2\rho+\tau}{\chi(m-1)}\Lambda\psi'' > 0$. The second term is clearly increasing in Λ. Therefore, overall the ratio ν_1'/ν_2' is increasing in Λ.

the H and M steady states disappear. We can show that

$$\underline{M} \to 1 + \frac{(\chi\overline{\Lambda} - \rho)(2\rho + \tau)}{\chi\left(\overline{\Lambda}(\chi - \tau) - \rho\right)}.$$

$$\overline{M} = \begin{cases} \frac{\chi+\rho}{\chi-\tau-\rho} & \text{if } \chi - \tau - \rho > 0 \\ +\infty & \text{if otherwise.} \end{cases}$$

Figure A-1: The M and L Steady States



Note: This figure shows how the M and L steady states are determined. For details of sample selection, see Appendix A.2.

In sum, when $m < \underline{M}$, there is only one H steady state.

When $\underline{M} \leq m \leq \overline{M}$, there are three steady states, H, M, and L.

$$\begin{aligned}
\alpha_H^* &= 1; \\
\alpha_M^* &\to \frac{\chi - \tau - \frac{2\rho+\tau}{m-1}}{\rho}; \qquad \text{(A-6)} \\
\alpha_L^* &\to \frac{\tau}{\chi\overline{\Lambda} - \rho}.
\end{aligned}$$

In the M steady state, $\Lambda_M^* < \overline{\Lambda}$ and $\omega_M^* = 1$. In the L steady state, $\Lambda_L^* = \overline{\Lambda}$ and $\omega_L^* > 1$.

When $m > \overline{M}$, there is only one L steady state.

The stability properties of the M and the L steady state are easily established. In the jargon of economy theory, a steady state is locally stable if, given an initial condition (in our case, an initial value for α_0) in the neighborhood of the steady state, there exists an equilibrium path converging to the steady state as $t \to \infty$. We know that in a neighborhood of either the M or the L steady state, the differential equation (A-4) must hold. In order to have $\lambda^f(t)$ converge to its steady state value $\lambda^{f*} = \frac{\rho+\tau}{\chi}$, we must necessarily have $\lambda^f(t) = \frac{\rho+\tau}{\chi}$ for all t.[11] Substituting $\lambda^f(t) = \frac{\rho+\tau}{\chi}$ into (A-3) and combine with $\Pi(t) = (m-1)(1-\alpha(t)\Lambda(t)$, we have

$$\begin{aligned}
\alpha(t) &= \frac{1}{\Lambda(t)}\left(1 - \frac{1\rho+\tau}{\chi(m-1)}\omega(\Lambda)\right) \\
&= \nu_1(\Lambda(t)).
\end{aligned}$$

In other words, starting from some $\alpha(0)$, a hypothetical convergence path coincides with the curve ν_1 in Figure A-1.

Recall the curve ν_2 in Figure A-1 describes the combination of α and Λ such that $\dot{\alpha} = 0$. This implies, starting from any $\alpha(t)$ below the ν_2 curve, we have $\dot{\alpha}(t) < 0$ and starting from an $\alpha(t)$ above the curve, we have $\dot{\alpha}(t) > 0$.

For any $\alpha \in (\alpha_L^*, \alpha_M^*)$, we have $\nu_1 > \nu_2$. It follows that, on the hypothetical converging trajectory, we must have $\dot{\alpha}(t) < 0$. Hence, starting from any $\alpha(0) \in (\alpha_L^*, \alpha_M^*)$, there exists a unique initial value $\Lambda(0)$ on ν_1 such that the equilibrium pair $(\alpha(t), \Lambda(t))$ travels southeast along the curve ν_1 and converges to the L steady state as $t \to \infty$. Similarly, pick any $\alpha(0) < \alpha_L^*$, we have $\dot{\alpha}(t) > 0$ along the trajectory, which implies that there exists a unique equilibrium pair $(\alpha(t), \Lambda(t))$ which travels northwest along ν_1 and converges to the L steady state. Finally, for any $\alpha(0) > \alpha_M^*$ we have $\nu_1 < \nu_2$. Therefore, any path starting and lying on ν_1 is characterized by $\dot{\alpha}(t) > 0$ for all t, which shows that there is no initial condition $\alpha(0)$ in the neighborhood of α_M^* for which we can find an equilibrium path converging to the M steady state. We then say that M is a source. L is a saddle to which we associate a unique converging path for any $\alpha(0) \in (\alpha_L^*, \alpha_M^*)$.

[11] Recall that necessarily $2\lambda^f - \tau/\chi > 0$. If $\lambda^f(t)$ increases to the steady state value, then $\lambda^f(t) > \frac{\rho+\tau}{\chi}$ and $\lambda^f(t)$ will increase without bound. If $\lambda^f(t)$ decreases to the steady state value, then $\lambda^f(t) > \frac{\rho+\tau}{\chi}$ and $\lambda^f(t)$ will decrease to zero. Either is a contradiction.

Clearly, if $m > \bar{M}$ then the only steady state is L, and for any initial condition $\alpha_0 \in (0, 1]$ the only equilibrium is the one associated with the unique path converging to the L steady state.

A.3 The Proof of Corollary 1

Firstly, note that aggregate R&D labor in the M and L can be expressed by

$$\alpha_i^* \Lambda_i^* = \frac{\tau \Lambda_i^*}{\tau + \chi(\Lambda_i^* - \lambda_i^{f*})}, \text{ for } i = M, L.$$

Since $\lambda_M^{f*} = \lambda_L^{f*}$ and $\Lambda_M^* < \Lambda_L^* = \overline{\Lambda}$, we conclude that $\alpha_M^* \Lambda_M^* < \alpha_L^* \Lambda_L^*$ under the assumption that $\tau < \chi - \rho < \chi$. Since $g_i^* = \log(\gamma) \alpha_i^* \Lambda_i^*$, we have $g_M^* > g_L^*$.

Secondly, comparing the aggregate production labor in the H and M steady states, (18) and (23), we find $L_H^* < L_M^*$ if and only if $m < \overline{M}$. Since $g_i^* = \log(\gamma)\chi(1 - L_i^*)$, we have $g_H^* > g_M^*$ when both exist.

Finally, we can easily solve out the growth rate in the L steady state: $g_L^* = \log(\gamma)\chi \frac{\tau\chi\overline{\Lambda}}{\chi\overline{\Lambda} - \rho}$, which is independent of m. Therefore, g_L^* is smaller than any growth rates in the M and H steady states for any $m \in (\underline{\mathrm{M}}, \overline{M})$.

B Dynamic Race with Endogenous Steps

Assume that the innovation technologies used by followers are linear and that innovation costs for followers are increasing in the follower's lag from the leader.[12] We will show that, under these assumptions, the model with an exogenous maximum distance of two steps is in fact the equilibrium result of a model where the maximum distance is endogenous. This conclusion holds for both the "step-by-step catch-up" and the "fast catch-up" versions of the model. For brevity, we present only the fast catch-up case.

To prove the claim, we need to consider only the problem of the follower, taking as given the value functions V_s^l of a leader s steps ahead, for $s = 1, 2, \ldots$. Also, to simplify the exposition, we can focus only on steady states. The assumption that innovation costs increase in the distance is modeled by assuming that a follower with lag $s > 1$ must first spend resources to close the gap to $s = 1$, and only then can try to leap-frog the leader. In the "fast catch-up case," followers can jump immediately from any state $s > 1$ to state $s = 1$ without having to retrace every intermediate step (as in a "step-by-step catch-up").

[12]Note that we make no assumption about the form of the innovation costs of the leader.

The value V_s^f of a follower $s \geq 2$ steps behind is given by the solution to

$$rV_s^f = \max_{\lambda_s^f \geq 0} -\chi_s \lambda_s^f + (\lambda_s^f + \tau_s)(V_1^f - V_s^f) + \lambda_s^l (V_{s+1}^f - V_s^f). \tag{B-1}$$

The value function of a follower $s = 1$ step behind solves the usual problem,

$$rV_1^f = \max_{\lambda_1^f \geq 0} -\lambda_1^f + \lambda_1^f (V_1^l - V_1^f) + \lambda_1^l (V_2^f - V_1^f). \tag{B-2}$$

Innovation costs χ_s are assumed to be increasing on the lag s, while the spillover intensity τ_s is assumed to be a decreasing sequence. We employ the normalizations $\chi_1 = 1$ and $\tau_1 = 0$.

For brevity we can appeal to an intuitive argument that, since innovation costs are increasing in the follower's lag, and spillover's intensities are decreasing, then $V_1^f \geq V_2^f \geq 0$, i.e. the follower is at least as well-off when he is one step behind the leader compared to when he is two steps behind.[13] Now, regardless of whether the condition $\lambda_1^f \geq 0$ is binding in the maximization of (B-2), and assuming that the optimal value of λ_1^f is finite, we have

$$V_1^f = \frac{\lambda_1^l}{r + \lambda_1^l} V_2^f.$$

Since $r > 0$ and $\lambda_1^l \geq 0$, the equation above and the inequalities $V_1^f \geq V_2^f \geq 0$ are satisfied if and only if

$$V_1^f = V_2^f = 0.$$

Substituting $V_1^f = V_2^f = 0$ in (B-1) for $s = 2$, the solution to the maximization gives optimal values $\lambda_2^f = 0$ and $V_3^f = 0$. Iterating the procedure for $s = 3, 4, \ldots$ yields

$$V_s^f = \lambda_s^f = 0, \forall s > 1.$$

This concludes the proof that, provided that the optimal λ_s^f is finite, followers never innovate when they are more than one step behind the leader.

[13] The result that the value to the follower decreases with the lag is standard in models of races (see for instance Hörner (2004)). For the sake of our demonstration, we can make the (incorrect) assumption that $V_2^f > V_1^f \geq 0$. Then, optimality of (B-1) for $s = 2$ requires that $\lambda_2^f = 0$. Moreover, since $V_2^f \geq 0$ and $\tau_2 > 0$, then $V_3^f - V_2^f > 0$ and thus $V_3^f - V_1^f > 0$. Iterating the argument for $s = 3, 4, \ldots$, we would conclude that $V_{s+1}^f - V_s^f > 0$ and λ_s^f for any $s > 1$. Hence, followers never innovate at stages $s > 1$.

C The Model with Constant Intertemporal Elasticity of Substitution Utility

The representative household solves the following problem:

$$\max \int_0^\infty e^{-\rho t} \frac{C(t)^{1-\sigma}}{1-\sigma} dt \tag{C-1}$$

$$\text{s.t.} \quad \int_0^\infty e^{-R(t)} E(t) dt \leq W(0), \tag{C-2}$$

where $R(t)$ is the compounded interest rate and $E(t)$ represents total spending at time t:

$$R(t) = \int_0^t r(\tau) d\tau,$$

$$E(t) = \int_{[0,1]} p_t(i) d_t(i) di.$$

The Cobb-Douglas form of the the consumption aggregate implies that the amount spent by the household on good i is the same across all products, giving

$$d_t(i) = \frac{E(t)}{p_t(i)}.$$

Therefore, we can write the consumption aggregate as

$$\log C(t) = \int_{[0,1]} \log \left(\frac{q_t(i)}{p_t(i)} E(t) \right) di = \int_{[0,1]} \log \left(\frac{q_t(i)}{p_t(i)} \right) di + \log E(t) \equiv \log Q(t) + \log E(t),$$

where $Q(t)$ is proportional to the aggregate quality index and $C(t) = Q(t)E(t)$.

We can rewrite the consumer's problem equivalently with a flow budget constraint, $\dot{a}(t) = r(t)a(t) + I(t) - E(t)$, where $a(t)$ is the stock of savings (wealth) at time t and $I(t)$ is the total income (labor income and profit from firms) at time t. We set up the current value Hamiltonian, $\mathcal{H}(E(t), a(t), \mu(t)) = \frac{(Q(t)E(t))^{1-\sigma}}{1-\sigma} + \mu(t)(r(t)a(t) + I(t) - E(t))$. The first order conditions, $\frac{\partial \mathcal{H}}{\partial E(t)} = 0$ and $\frac{\partial \mathcal{H}}{\partial a(t)} = \rho\mu(t) - \dot{\mu}(t)$, imply

$$\begin{aligned} r(t) &= \rho + \sigma \frac{\dot{E}(t)}{E(t)} + (\sigma - 1) \frac{\dot{Q}(t)}{Q(t)} \\ &= \rho + \sigma \frac{\dot{E}(t)}{E(t)} + (\sigma - 1) \log(\gamma) \dot{S}(t), \end{aligned}$$

which is (26) in the paper. Also note that in the special case of log period utility ($\sigma = 1$), we obtain the familiar $\frac{\dot{E}(t)}{E(t)} = r(t) - \rho$.

Let's maintain all the parametric assumptions made in the baseline model and suppose $m \in \left(\underline{M}, \overline{M}\right)$ so three steady states exist under the baseline assumption $\sigma = 1$. We characterize the structure of the steady states in this environment when σ deviates from 1.

The H steady state. The highest feasible steady state value for α^* is one, since in this case $\lambda^{l*} = 0$. Let's first assume that around the H steady state we have $\lambda^{f*} > 0$. As usual, a steady state with high extensive margin will be associated with a low intensive margin Λ^*. By taking $\bar{\Lambda}$ large enough, we can make sure that in a neighborhood of the steady state $\Lambda^*(t) = \lambda^{f*}(t) < \bar{\Lambda}$, giving $\omega(t) = 1$. Hence, the first order condition for λ^f in a neighborhood of a H steady state implies that

$$V_1^l(t) = \frac{1}{\chi}.$$

The condition above implies that, in a neighborhood of the H steady state, $\dot{V}_1^l = 0$. Since $\lambda^{l*}(t) = 0$, a straightforward substitution in the definition of V_1^l gives

$$\frac{r(t)}{\chi} + \lambda^f(t) = \Pi(t). \tag{C-3}$$

Combining the above equation with the facts that $\Pi(t) = (m-1)L(t)$, $\lambda^f(t) = (1-L(t))/\alpha(t)$ and (26), we obtain

$$\frac{\rho}{\chi} + \frac{\sigma}{\chi}\frac{\dot{L}}{L} + (\sigma - 1)\log(\gamma)(1 - L) + \frac{1 - L}{\alpha} = (m - 1)L. \tag{C-4}$$

Equation (C-4) defines the evolution of the economy around the H steady state, together with the condition

$$\dot{\alpha} = \tau(1 - \alpha). \tag{C-5}$$

The H steady state is then characterized by

$$\alpha_H^* = 1;$$
$$L_H^* = \frac{1 + \rho/\chi + (\sigma - 1)\log(\gamma)}{m + (\sigma - 1)\log(\gamma)};$$
$$\lambda_H^{f*} = \frac{m - 1 - \rho/\chi}{m + (\sigma - 1)\log(\gamma)};$$
$$g_H^* = \log(\gamma)\frac{\chi(m - 1) - \rho}{m + (\sigma - 1)\log(\gamma)}.$$

Linearizing the system (15), (16) we can show that the H steady state is a saddle. We maintain the assumption from the baseline model that $m > 1 + \rho/\chi$ to have a non-degenerate

H steady state. Moreover, $\lambda_H^{f*} > 0$ and $L_H^* > 0$ jointly requires

$$\sigma > 1 - \frac{1 + \frac{\rho}{\chi}}{\log(\gamma)} \equiv \underline{\sigma}_{H1}.$$

For the H steady state to exist, we in addition require $V_2^{l*} < \frac{2}{\chi}$ so that leaders indeed do not have incentive to innovate. This gives us

$$\begin{aligned}
V_2^{l*} &= \frac{(m-1)L_H^* + \tau V_1^{l*}}{\rho + (\sigma - 1)g_H^* + \tau} \\
&= \frac{(m-1)\chi\left[1 + \rho/\chi + (\sigma - 1)\log(\gamma)\right] + \tau(m + (\sigma - 1)\log(\gamma))}{(\rho + \tau)\chi(m + (\sigma - 1)\log(\gamma)) + (\sigma - 1)\log(\gamma)(\chi(m-1) - \rho)\chi} < \frac{2}{\chi} \\
\Rightarrow \quad \sigma &> 1 - \frac{\left(\frac{\rho+\chi}{\chi-\rho-\tau} - m\right)(\chi - \rho - \tau)}{\log(\gamma)(\chi(m-1) + \tau)} \\
&= 1 - \frac{\left(\overline{M} - m\right)(\chi - \rho - \tau)}{\log(\gamma)(\chi(m-1) + \tau)} \equiv \underline{\sigma}_{H2}.
\end{aligned}$$

We maintain the assumption from the baseline that $\chi > \rho + \tau$. For the set of parameters under which the H steady state exists in the baseline model (i.e. $m < \overline{M}$), the H steady state exists in this extended model as long as σ is greater than $\underline{\sigma}_{H2}$, which is a number less than 1.

To compare $\overline{\sigma}_{H1}$ and $\overline{\sigma}_{H2}$, we first note that

$$\begin{aligned}
&\frac{\rho + \chi}{\chi} - \frac{\left(\overline{M} - m\right)(\chi - \rho - \tau)}{\chi(m-1) + \tau} \\
=&\frac{(2\chi - \tau)\chi\left(m - 1 - \frac{\rho}{\chi}\right)}{\chi\left(\chi(m-1) + \tau\right)},
\end{aligned}$$

which has the same sign as $m - 1 - \frac{\rho}{\chi}$. Recall that $m > \underline{M} = 1 + \frac{(2\rho+\tau)(\chi\overline{\Lambda}-\rho)}{\chi((\chi-\tau)\overline{\Lambda}-\rho)}$ and $\underline{M}$ is decreasing in $\overline{\Lambda}$. This implies

$$m > \underline{M} > \lim_{\overline{\Lambda} \to +\infty} \underline{M} = 1 + \frac{2\rho + \tau}{\chi - \tau} > 1 + \frac{\rho}{\chi}.$$

Therefore, $m - 1 - \frac{\rho}{\chi} > 0$, $\frac{\rho+\chi}{\chi} > \frac{\left(\overline{M}-m\right)(\chi-\rho-\tau)}{\chi(m-1)+\tau}$, and in turn

$$\underline{\sigma}_{H1} < \underline{\sigma}_{H2}.$$

Define $\underline{\sigma}_H = \underline{\sigma}_{H2}$. The H steady state exists as long as $\sigma \geq \underline{\sigma}_H$. At $\underline{\sigma}_H$, leaders are indifferent between innovating or not. For a σ that is infinitesimally smaller than $\underline{\sigma}_H$, leaders will have strictly prefer to innovate.

The M and L steady states In the M and L steady states, both leaders and followers innovate at the contestable state. The first order conditions for λ^l and λ^f give $V_2^l(t) = 2\omega(t)/\chi = 2V_1^l(t)$. Substituting these conditions into the value functions and after appropriate calculations we obtain the two equations:

$$\begin{cases} \Pi = (2\lambda^f - \frac{\tau}{\chi})\omega \\ \dot{\omega} = (r + \tau - \chi\lambda^f)\omega. \end{cases} \tag{C-6}$$

From the first equation in (C-6) and $\Pi = (m-1)L$ we can solve out L:

$$L = \frac{2\lambda^f - \tau/\chi}{m-1}\omega,$$

which, together with the labor market clearing condition, implies

$$1 - \frac{2\lambda^f - \tau/\chi}{m-1}\omega = \alpha(\lambda^f + \lambda^l). \tag{C-7}$$

From the second equation in (C-6), in the steady state $r + \tau - \chi\lambda^f = 0$. Combined with the Euler equation derived at the beginning of this section, we have

$$\rho + (\sigma - 1)g + \tau - \chi\lambda^f = 0. \tag{C-8}$$

The difference between an M and a L steady state is that in an M steady state, the skilled labor supply does not bind ($\lambda^l + \lambda^f < \overline{\Lambda}$) and the skill premium is one ($\omega = 1$), whereas in a L steady state, the opposite is true: $\lambda^l + \lambda^f = \overline{\Lambda}$ and $\omega > 1$.

Then we can use four equations to characterize an M steady state

$$\begin{cases} \alpha = \frac{\tau}{\tau + \chi\lambda^l} \\ g = \log(\gamma)\chi\alpha(\lambda^f + \lambda^l) \\ \rho + (\sigma - 1)g + \tau - \chi\lambda^f = 0 \\ 1 - \frac{2\lambda^f - \tau/\chi}{m-1} = \alpha(\lambda^f + \lambda^l) \end{cases}. \tag{C-9}$$

The first equation becomes from the evolution of the extensive margin $\alpha(t)$. The second equation is the definition of the growth rate. The third equation is (C-8). The last equation is (C-7), where $\omega = 1$ in an M steady state. From these four equations, we can solve for the M steady state endogenous variables: α_M^*, g_M^*, λ_M^{f*} and λ_M^{l*}.

We can use another set of four equations to characterize a L steady state

$$\begin{cases} \alpha = \frac{\tau}{\tau + \chi(\overline{\Lambda} - \lambda^f)} \\ g = \log(\gamma)\chi\alpha\overline{\Lambda} \\ \rho + (\sigma - 1)g + \tau - \chi\lambda^f = 0 \\ 1 - \frac{2\lambda^f - \tau/\chi}{m-1}\omega = \alpha\overline{\Lambda} \end{cases}. \tag{C-10}$$

From these four equations, we can solve for the L steady state endogenous variables: α_L^*, g_L^*, λ_L^{f*} and ω_L^*.

Let's focus on the M steady state first. Combining the second and fourth equation in (C-9), we have one equation that links g to λ^f:

$$g = \log(\gamma)\chi\frac{m-1-2\lambda^f+\tau/\chi}{m-1}. \tag{C-11}$$

Together with the third equation in (C-9), we can solve out the M steady state explicitly

$$\begin{aligned}
\alpha_M^* &= \frac{(\chi-\tau)(m-1)-(2\rho+\tau)-2\tau(\sigma-1)\log(\gamma)}{\rho(m-1)-[\tau-\chi(m-1)](\sigma-1)\log(\gamma)};\\
L_M^* &= \frac{2\rho+\tau+2\chi(\sigma-1)\log(\gamma)}{\chi(m-1)+2\chi(\sigma-1)\log(\gamma)};\\
\lambda_M^{f*} &= \frac{(\sigma-1)\log(\gamma)\chi[m-1+\tau/\chi]+(m-1)(\rho+\tau)}{[m-1+2(\sigma-1)\log(\gamma)]\chi};\\
g_M^* &= \log(\gamma)\frac{\chi(m-1)-2\rho-\tau}{m-1+2(\sigma-1)\log(\gamma)}.
\end{aligned}$$

The λ_M^{l*} is implied in the last equation of (C-9). Rearranging terms,

$$\lambda_M^{l*} = \frac{\tau\left((m+1)\lambda_M^{f*}-m+1-\tau/\chi\right)}{(m-1)(\chi-\tau)+\tau-2\chi\lambda_M^{f*}}, \tag{C-12}$$

which is increasing in λ_M^{f*}. All these endogenous variables are well-defined when $\sigma=1$. Let's differentiate λ_M^{f*} with respect to $\sigma-1$.

$$\frac{d\lambda_M^{f*}}{d(\sigma-1)} = \frac{\log(\gamma)(m-1)\chi\left[(m-1)\chi-2\rho-\tau\right]}{[m-1+2(\sigma-1)\log(\gamma)]^2\chi^2}.$$

Recall that $m > \underline{M} = 1+\frac{(2\rho+\tau)(\chi\overline{\Lambda}-\rho)}{\chi((\chi-\tau)\overline{\Lambda}-\rho)} > 1+\frac{2\rho+\tau}{\chi}$. Hence,

$$\frac{d\lambda_M^{f*}}{d(\sigma-1)} > 0.$$

This means, as σ decreases below 1, both λ_M^{f*} and λ_M^{l*} will decrease until λ_M^{l*} becomes zero, at which point the M steady state coincides with the H steady state when leaders are indifferent between innovating and not innovating. To see this point, when $\lambda_M^{l*}=0$, from (C-12), λ_M^{f*} becomes

$$\lambda_M^{f*} = \frac{m-1+\tau/\chi}{m+1}.$$

Evaluate the λ_H^{f*} at $\sigma = \underline{\sigma}_H$:

$$\lambda_H^{f*} = \frac{m - 1 - \rho/\chi}{m - \frac{(\overline{M} - m)(\chi - \rho - \tau)}{\chi(m-1)+\tau}} = \frac{m - 1 + \tau/\chi}{m+1} = \lambda_M^{f*}.$$

This also means, As σ rises above 1, both λ_M^{f*} and λ_M^{l*} will increase until the sum hits the fixed supply: $\lambda_M^{f*} + \lambda_M^{l*} = \overline{\Lambda}$. At this point, as we will show below, the M steady state coincides with the L steady state where $\omega = 1$.

Let's focus on the L steady state now. Combining the first two equations in (C-10) and cancelling out α, we have

$$\frac{g}{\log(\gamma)\chi\overline{\Lambda}} = \frac{\tau}{\tau + \chi(\overline{\Lambda} - \lambda^f)}.$$

Combining the above with the third equation in (C-10), we can infer the L steady state λ_L^{f*} from

$$\chi^2 \lambda_L^{f*2} - (\rho + 2\tau + \chi\overline{\Lambda})\chi\lambda_L^{f*} + (\rho + \tau)(\tau + \chi\overline{\Lambda}) + (\sigma - 1)\log(\gamma)\tau\chi\overline{\Lambda} = 0. \tag{C-13}$$

Under our assumption of $m \in \left(\underline{M}, \overline{M}\right)$, we know when $\sigma = 1$ there exists a well-defined L steady state. When $\sigma = 1$, the above quadratic has two roots: $\lambda^f = \frac{\rho+\tau}{\chi}$ and $\lambda^f = \overline{\Lambda} + \frac{\tau}{\chi}$ (omitted because it is greater than $\overline{\Lambda}$). The smaller root is the R&D intensity of the followers in the L steady state in the baseline model, λ_L^{f*}, and we also know in that steady state $\omega_L^* > 1$. As σ increases above 1, the quadratic function shifts up and the smaller root, λ_L^{f*}, increases, which in turn implies that α_L^* increases (see the first equation of (C-10)). Now from the fourth equation in (C-10), we deduce that the steady state ω_L^* must decrease. Therefore, as σ increases, the smaller root to (C-13) defines the L steady state level of λ^f until the implied ω_L^* decreases to 1, at which point the L steady state coincides with the M steady state where the constraint on skilled labor supply becomes just binding. To see this point, note how the solution to (C-9) when $\lambda^f + \lambda^l = \overline{\Lambda}$ must also solve (C-10) when $\omega = 1$ and vice versa.

Finally, we show that the larger root of this quadratic equation (C-13) can never be a L steady state. Since σ only shifts the quadratic function up and down, the larger root will always be strictly larger than the λ_L^{f*} when L and M steady states coincide as we discuss above. Suppose the larger root, λ_2^f occurs in a L steady state. Then in that steady state, the extensive margin α must be larger than the extensive margin when L and M steady states coincide. This also means, ω in that steady state much be strictly smaller than the skill premium when L and M steady states coincide, which we know is 1. This contradicts the definition of a L steady state.

Let $\overline{\sigma}_L$ be the σ at which the L and M steady states coincide and let $\underline{\sigma}_L$ be the σ when the smaller root of the quadratic equation is $\frac{\tau}{2\chi}$ and $\underline{\sigma}_L < 1$. We have shown that for $\sigma \in (\underline{\sigma}_L, \overline{\sigma}_L)$, the L steady state exists. λ_L^{f*} is given by the smaller root of equation (C-13) and the other steady state variables can be derived by

$$\alpha_L^* = \frac{\tau}{\tau + \chi(\overline{\Lambda} - \lambda_L^{f*})};$$
$$g_L^* = \alpha_L^* \log(\gamma)\chi\overline{\Lambda};$$
$$L_L^* = 1 - \alpha_L^*\overline{\Lambda}.$$

We next show that $\underline{\sigma}_L < \underline{\sigma}_H$, such that the L steady state is defined whenever the H steady state is defined and $\sigma < 1$. Substituting λ^f with $\frac{\tau}{2\chi}$ in (C-13), we can rearrange to obtain

$$(1 - \underline{\sigma}_L)\log(\gamma) = \frac{(2\rho + \tau)(2\chi\overline{\Lambda} + \tau)}{4\tau\chi\overline{\Lambda}}.$$

Recall $\underline{\sigma}_H = \underline{\sigma}_{H2}$ and

$$(1 - \underline{\sigma}_H)\log(\gamma) = \frac{\chi + \rho - m(\chi - \rho - \tau)}{\chi(m-1) + \tau}.$$

We can derive the following inequalities

$$\begin{aligned}(1 - \underline{\sigma}_H)\log(\gamma) &< \frac{(2\rho + \tau)\rho(\chi\overline{\Lambda} - \tau - \rho)}{2\chi(\rho + \tau)(\chi\overline{\Lambda} - \rho) - \tau^2\overline{\Lambda}\chi} \\ &= \frac{2\rho + \tau}{\chi\overline{\Lambda}} \frac{\rho(\chi\overline{\Lambda} - \tau - \rho)}{2(\rho + \tau)\left(\chi - \frac{\rho}{\overline{\Lambda}}\right) - \tau^2} \\ &< \frac{2\rho + \tau}{\chi\overline{\Lambda}} \frac{\rho(\chi\overline{\Lambda} - \tau - \rho)}{\tau(2\chi - \tau)}.\end{aligned}$$

The first inequality is obtained by replacing m by $\underline{M}$ since $(1 - \underline{\sigma}_H)\log(\gamma)$ decreases in m and $m > \underline{M}$. The second inequality is obtained by replacing $\left(\chi - \frac{\rho}{\overline{\Lambda}}\right)$ on the denominator by $\frac{\chi\tau}{\rho+\tau}$ since $\overline{\Lambda} > \frac{\rho+\tau}{\chi}$. Now, we have

$$\begin{aligned}& (1 - \underline{\sigma}_H)\log(\gamma) < (1 - \underline{\sigma}_L)\log(\gamma) \\ \Leftarrow\quad & \frac{\rho(\chi\overline{\Lambda} - \tau - \rho)}{2\chi - \tau} < \frac{2\chi\overline{\Lambda} + \tau}{4} \\ \Leftrightarrow\quad & 4\rho(\chi\overline{\Lambda} - \tau - \rho) < (2\chi - \tau)(2\chi\overline{\Lambda} + \tau) \\ \Leftrightarrow\quad & 2\chi(2\rho - 2\chi + \tau)\overline{\Lambda} < (2\chi - \tau)\tau + 4\rho(\rho + \tau),\end{aligned}$$

which is always true. Because we maintain the assumption that $\chi > \rho + \tau$, the left hand side of the above inequality is negative where as the right hand side is positive. Hence, we conclude

$$\underline{\sigma}_L < \underline{\sigma}_H < 1.$$

This means, the L steady state is always defined for any $\sigma < 1$ under which the H steady state is also defined.

We summarize the discussions above into the following proposition, which also characterizes the stability properties of the different steady states. Maintain the assumption that $m \in \left(\underline{M}, \overline{M}\right)$.

Proposition 2. *There exist $\underline{\sigma}_H$ and $\overline{\sigma}_L$ such that $\underline{\sigma}_H < 1 < \overline{\sigma}_L$. For $\sigma \in (\underline{\sigma}_H, \overline{\sigma}_L)$, the economy has three steady states, H, M, and L. The H and L steady states are saddle path stable, while the M steady state is unstable. For $1 \leq \sigma < \overline{\sigma}_L$, the three steady states can be ranked by the aggregate growth rates, $g_H^* > g_M^* > g_L^*$.*

We show how aggregate growth is ordered in the three steady states as described in Proposition 2. First, we show that the growth rate in the H steady state is always higher than that in the M steady state. Since $g_i^* = \log(\gamma)\chi(1 - L_i^*)$ for $i = M, H$, it suffices to show that $L_H^* < L_M^*$.

$$
\begin{aligned}
& L_H^* < L_M^* \\
\Leftrightarrow \quad & \frac{1 + \rho/\chi + (\sigma - 1)\log(\gamma)}{m + (\sigma - 1)\log(\gamma)} < \frac{2\rho + \tau + 2\chi(\sigma - 1)\log(\gamma)}{\chi(m - 1) + 2\chi(\sigma - 1)\log(\gamma)} \\
\Leftrightarrow \quad & (\sigma - 1)\log(\gamma) > -\frac{\chi + \rho - m(\chi - \rho - \tau)}{\chi(m - 1) + \tau} \\
\Leftrightarrow \quad & \sigma > \underline{\sigma}_H,
\end{aligned}
$$

an assumption made in Proposition 2. Hence, we have $g_H^* > g_M^*$.

Next we order g_M^* and g_L^*. We first introduce the following Lemma which can be proved by contracditions.

Lemma 1. *In the M and L steady states, we have $\lambda_M^{f*} \geq \lambda_L^{f*}$, if $\sigma \geq 1$.*

Proof. The case of $\sigma = 1$ is discussed in the baseline model, in which case $\lambda_M^{f*} = \lambda_L^{f*}$. Suppose $\sigma > 1$ and we prove by contradiction. Suppose $\lambda_M^{f*} \leq \lambda_L^{f*}$. Since $\lambda_i^{f*} = \frac{r_i^* + \tau}{\chi}$, for $i = M, L$, it implies that $r_M^* \leq r_L^*$. Since $r_i^* = \rho + (\sigma - 1)g_i^*$, it implies that $g_M^* \leq g_L^*$. Since $g_i^* = \log(\gamma)\chi(1 - L_i^*)$, we have $L_M^* \geq L_L^*$. On the other hand, from the first equation of (C-6), it must be true that

$$
L_M^* = \frac{2\lambda_M^{f*} - \frac{\tau}{\chi}}{m - 1} < \frac{2\lambda_L^{f*} - \frac{\tau}{\chi}}{m - 1} < \frac{2\lambda_L^{f*} - \frac{\tau}{\chi}}{m - 1}\omega_L = L_L^*,
$$

since $\omega_L^* > 1$. We reach a contradiction. □

This, together with the equation $\rho + (\sigma - 1)g_i^* + \tau - \chi\lambda_i^{f*} = 0$ for $i = M, L$, implies that $g_M^* > g_L^*$ as long as $\sigma > 1$. This concludes the proof for $g_H^* > g_M^* > g_L^*$ for $1 \leq \sigma < \overline{\sigma}_L$. The proof of the local stability properties of the steady states is available upon request.

D The Model with A Quadratic Cost of Innovation

Replace the linear cost of innovation in the baseline model with the following quadratic cost. In order to achieve an arrival rate of innovation of λ, the firm needs to employ $C(\lambda)$ skilled labor (at a wage normalized to 1):

$$C_i(\lambda) = \chi_i \lambda + \frac{1}{2}\theta_i \lambda^2,$$

where χ_i and θ_i are parameters of the cost function for leaders ($i = 1$) and followers ($i = 2$). The value functions of leaders and followers are given as follows.

$$rV_2^l = \Pi + \tau(V_1^l - V_2^l) + \dot{V}_2^l \tag{D-1}$$

$$rV_2^f = \tau(V_1^f - V_2^f) + \dot{V}_2^f \tag{D-2}$$

$$rV_1^l = \max_{\lambda^l \geq 0} \Pi - \chi_1 \lambda^l - \frac{1}{2}\theta_1 \lambda^{l2} + \lambda^l(V_2^l - V_1^l) + \lambda^f(V_1^f - V_1^l) + \dot{V}_1^l \tag{D-3}$$

$$rV_1^f = \max_{\lambda^f \geq 0} -\chi_2 \lambda^f - \frac{1}{2}\theta_2 \lambda^{f2} + \lambda^f(V_1^l - V_1^f) + \lambda^l(V_2^f - V_1^f) + \dot{V}_1^f \tag{D-4}$$

The FOCs imply

$$V_2^l - V_1^l = \chi_1 + \theta_1 \lambda^l \tag{D-5}$$

$$V_1^l - V_1^f = \chi_2 + \theta_2 \lambda^f. \tag{D-6}$$

D.1 Both Leaders and Followers Innovating

Focus on the steady states where both leaders and followers innovate. Subtracting (D-3) from (D-1) and rearranging,

$$(r + \tau + \lambda^l)(V_2^l - V_1^l) = \chi_1 \lambda^l + \frac{1}{2}\theta_1 \lambda^{l2} + \lambda^f(V_1^l - V_1^f),$$

where $V_2^l - V_1^l$ is given by (D-5) and $V_1^l - V_1^f$ is given by (D-6), and $r = \rho$ in a steady state. This implies the first equation that involves λ^f and λ^l:

$$(\rho + \tau)(\chi_1 + \theta_1 \lambda^l) + \frac{1}{2}\theta_1 \lambda^{l2} = \lambda^f(\chi_2 + \theta_2 \lambda^f). \tag{D-7}$$

Subtracting (D-4) from (D-2) and rearranging,

$$V_2^f - V_1^f = \frac{\chi_2 \lambda^f + \frac{1}{2}\theta_2 \lambda^{f2} - \lambda^f(V_1^l - V_1^f)}{r + \tau + \lambda^l} = \frac{-\frac{1}{2}\theta_2 \lambda^{f2}}{r + \tau + \lambda^l}, \tag{D-8}$$

where the last equality follows from substituting $V_1^l - V_1^f$ by (D-6).

Subtracting (D-4) from (D-3) and rearranging,

$$(r+2\lambda^f)(V_1^l - V_1^f) = \Pi^* - \chi_1\lambda^l - \frac{1}{2}\theta_1\lambda^{l2} + \lambda^l(V_2^l - V_1^l) + \chi_2\lambda^f + \frac{1}{2}\theta_2\lambda^{f2} - \lambda^l(V_2^f - V_1^f),$$

where $V_1^l - V_1^f$ is given by (D-6), $V_2^l - V_1^l$ is given by (D-5), $V_2^f - V_1^f$ is given by (D-8). Substituting these terms in the above equation, we have

$$(r+2\lambda^f)(\chi_2 + \theta_2\lambda^f) = \Pi^* + \frac{1}{2}\theta_1\lambda^{l2} + \lambda^f \frac{(r+\tau)(\chi_2 + \frac{1}{2}\theta_2\lambda^f) + \lambda^l(\chi_2 + \theta_2\lambda^f)}{r+\tau+\lambda^l}. \quad \text{(D-9)}$$

Note that in a steady state where both leaders and followers innovate, the extensive margin is given by

$$\alpha^* = \frac{\tau}{\tau + \lambda^l}.$$

Then, profit in the steady state becomes

$$\begin{aligned}\Pi^* &= (m-1)\left[1 - \alpha^*\left(\chi_1\lambda^l + \frac{1}{2}\theta_1\lambda^{l2} + \chi_2\lambda^f + \frac{1}{2}\theta_2\lambda^{f2}\right)\right] \\ &= (m-1)\left[1 - \frac{\tau}{\tau+\lambda^l}\left(\chi_1\lambda^l + \frac{1}{2}\theta_1\lambda^{l2} + \chi_2\lambda^f + \frac{1}{2}\theta_2\lambda^{f2}\right)\right],\end{aligned}$$

which we can plug in (D-9) together with $r = \rho$ to obtain

$$\begin{aligned}&\theta_2\left[\frac{1}{2}\left(\frac{\rho+\tau+2\lambda^l}{\rho+\tau+\lambda^l} - \frac{(m-1)\tau}{\tau+\lambda^l}\right) - 2\right]\lambda^{f2} - \left(\frac{(m-1)\chi_2\tau}{\tau+\lambda^l} + \chi_2\rho\theta_2\right)\lambda^f \\ &\qquad + (m-1)\left[1 - \frac{\tau}{\tau+\lambda^l}\left(\chi_1\lambda^l + \frac{1}{2}\theta_1\lambda^{l2}\right)\right] + \frac{1}{2}\theta_1\lambda^{l2} - \rho\chi_2 = 0.\end{aligned} \quad \text{(D-10)}$$

Equations (D-7) and (D-10) form a system of equations, from which we can solve for λ^l and λ^f, which give us the steady state λ^{l*} and λ^{f*}.

D.2 Only Followers Innovating

Now consider the steady state, where only followers innovate. In this steady state, $\alpha^* = 1$ and $\lambda^{l*} = 0$.

The value functions, (D-3) and (D-4), at the steady state become

$$\begin{aligned} rV_1^l &= \Pi + \lambda^f(V_1^f - V_1^l) \\ rV_1^f &= -\chi_2\lambda^f - \frac{1}{2}\theta_2\lambda^{f2} + \lambda^f(V_1^l - V_1^f).\end{aligned}$$

Taking the difference of the above two equations, we have

$$(r + 2\lambda^f)(V_1^l - V_1^f) = \Pi + \chi_2\lambda^f + \frac{1}{2}\theta_2\lambda^{f2}. \tag{D-11}$$

Note that the profit is given by

$$\Pi = (m-1)L = (m-1)\left(1 - \chi_2\lambda^f - \frac{1}{2}\theta_2\lambda^{f2}\right). \tag{D-12}$$

Plugging (D-6) and (D-12) in (D-11) and replace r by the steady state value ρ, we have

$$\begin{aligned}
&(\rho + 2\lambda^f)(\chi_2 + \theta_2\lambda^f) = (m-1)\left(1 - \chi_2\lambda^f - \frac{1}{2}\theta_2\lambda^{f2}\right) \\
\Rightarrow &\theta_2\left(\frac{1}{2}m + 1\right)\lambda^{f2} + (\rho\theta_2 + m\chi_2)\lambda^f + \rho\chi_2 - m + 1 = 0.
\end{aligned}$$

from which we can solve for the steady state value for λ^f, λ^{f*}.

From the corner solution for λ^f, we can infer that

$$V_2^l - V_1^l < \chi_1.$$

Taking the difference of V_2^l and V_1^l, we have

$$V_2^l - V_1^l = \frac{\lambda^{f*}}{\rho + \tau}(V_1^l - V_1^f).$$

From (D-11), we derive

$$\begin{aligned}
V_2^l - V_1^l &= \frac{\lambda^{f*}}{\rho + \tau}\frac{\Pi + \chi_2\lambda^{f*} + \frac{1}{2}\theta_2\lambda^{f*2}}{\rho + 2\lambda^{f*}} \\
&= \frac{\lambda^{f*}}{\rho + \tau}\frac{(m-1) - (m-2)\left(\chi_2\lambda^{f*} + \frac{1}{2}\theta_2\lambda^{f*2}\right)}{\rho + 2\lambda^{f*}} < \chi_1.
\end{aligned}$$

This is the condition for the existence of the steady state where leaders indeed do not innovate.

E Details of the Empirical Exercise

E.1 Data, Sample Selection and Summary Statistics

We use the Fundamental Annual Table of Compustat - Capital IQ from Standard & Poor's from 1962 to 2018. We keep all US firms, drop financial firms (whose SIC codes are between 6000 and 6999), drop utilities firms (whose SIC codes are between 4900 and 4999) and drop non classified or mostly government entities (whose SIC codes are above 9000). We keep firm-year observations with non-missing sales revenue and operating income. We also convert all nominal monetary values to real values (2005 dollar) using the CPI downloaded from the St. Louis Fed. In the analysis, we use the Fama-French industry classification that gives us 49 industries. Table E-1 summarizes the key variables used in the paper. As expected, all of the number of employees, sales revenue, and R&D expenses are highly right-skewed, with the means much larger than the median values.

To understand the prevalence of leader innovating, we define an industry leader as the firm that has the largest market share in one of the 49 Fama-French industry in a given year. We then calculate the share of industry total R&D that is conducted by the leaders for each of the 49 industries averaged over time. The results are reported in Figure E-1. To give a general impression, we aggregate leaders' shares of R&D to nine coarse industry group and report the results in Panel (a) of the figure. For the over economy, industry leaders on average conduct 23% of R&D whereas the average market share is 20%. In Panel (b), we separately report the share of R&D conducted by leaders for each of the top 10 R&D intensive Fama-French industries. The R&D intensity is defined as the average ratio of R&D expense to sales revenue across years. The intensities are given in the parenthesis next to the industry names. For example, the industry with the highest R&D intensity is the pharmaceutical industry, where 14.4% of sales revenue goes into R&D on average. Among the R&D intensive industries, leaders conduct from 12% (in pharmaceutical industry) to 81% (in agriculture industry) of industry total R&D, averaging at 20%.

We conclude that industry leaders are responsible for a sizable amount of R&D in the entire economy.

Table E-1: Summary Statistics, Compustat 1962-2018

	no. of obs	mean	st. dev.	p1	p25	p50	p75	p99
Number of employees (thousands)	236,293	6.90	32.86	0.00	0.12	0.70	3.50	103.70
Sales revenue (millions)	260,407	1710.00	9161.32	0.00	23.72	142.48	706.40	26699.15
R&D expenses (millions)	138,473	62.46	441.45	0.00	0.17	2.90	17.34	1099.94
Year	260,407	1992	14.25	1963	1981	1993	2003	2017

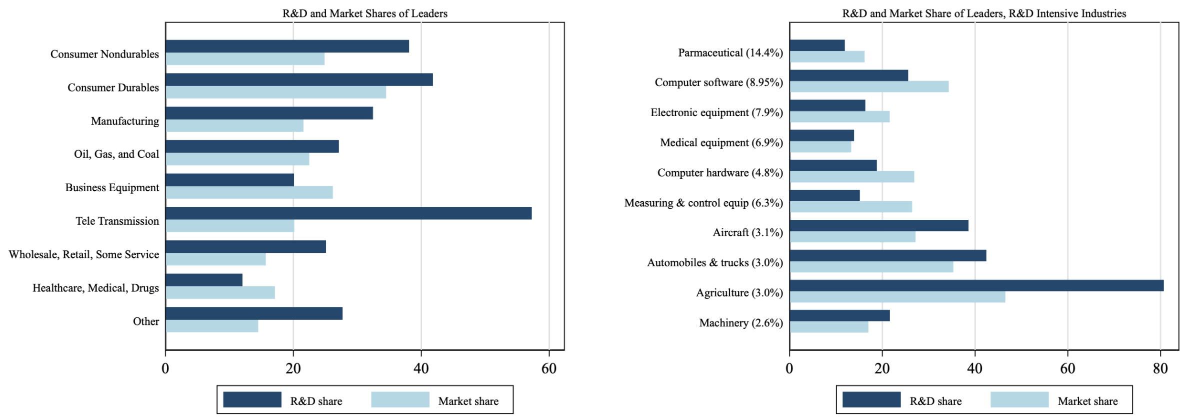
Note: This table provides the summary statistics of main variables that we use in the analysis from Compustat 1962-2018. For details of sample selection, see Appendix E.1.

E.2 Estimation Results

We select a panel of firms that have been an industry leader in at least one year in our sample. We run four regressions. We further keep firms that have at least 10 years of data. In effect, we drop one firm, Brightview Holding in Agriculture, which is in our sample only from 2016 to 2018. We keep the history of these firms until the last year they are observed to be leaders. In other words, we discard the behavior of the firm when we do not know if the behavior leads to an eventual leadership. When we take the first difference of log R&D

Figure E-1: Share of R&D Conducted by Industry Leaders

(a) All Industries (b) Top 10 R&D Intensive Industries



Note: This figure shows the percentage of R&D conducted by industry leaders, by coarse industry groups and by industry for a subsample of R&D intensive industries. For details of data construction, see Appendix E.1.

expenses to obtain R&D growth, we trim the top and bottom 1% of the growth. This gives us a sample of histories of R&D and market shares of eventual leaders, on which we run the following two regressions:

$$\log RnD_{it} = \alpha_0 MarketShare_{it} + \alpha_1 MarketShare_{it}^2 + \text{time f.e.} + \text{firm f.e.} \tag{E-1}$$
$$\Delta \log RnD_{it} = \beta MarketShare_{it} + \text{industryXyear f.e..} \tag{E-2}$$

Controlling for firm fixed effects and time fixed effects, regression (E-1) traces out how the level of R&D expenditures (RnD_{it}) evolves over an eventual leader's lifecycle as the firm sets on its course of capturing a larger and larger market share ($MarketShare_{it}$). Taking the first difference, regression (E-2) traces out how the growth rate of R&D relates to the market share.

Further conditioning on the behavior of leaders during leadership, we regress the R&D level and growth on the age of the leadership, i.e. the number of years into a continuous leadership ($Duration_{it}$):

$$\log RnD_{it} = \gamma_0 Duration_{it} + \gamma_1 Duration_{it}^2 + \text{time f.e.} + \text{firm f.e.} \tag{E-3}$$
$$\Delta \log RnD_{it} = \delta Duration_{it} + \text{time f.e.} + \text{industry f.e..} \tag{E-4}$$

The results are found in Table E-2, where each column corresponds to each of the above four regressions, (E-1) to (E-4). It is clear that over an eventual leader's lifecycle, the relationship between R&D and market share is an inverted U. Equally clear is the relationship between R&D and age of the leadership conditioning on being a leader, which also exhibits an inverted U shape.

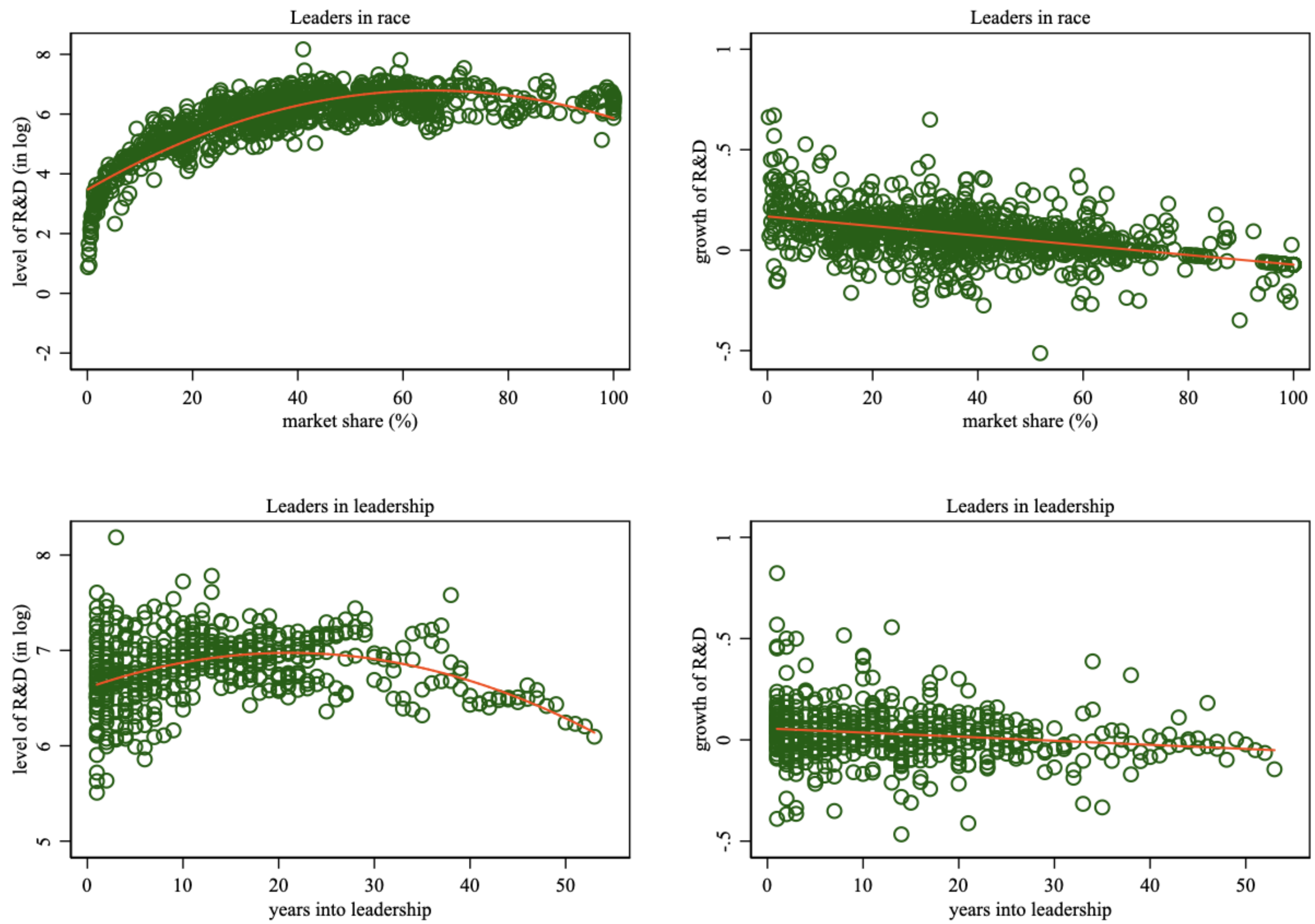
If we plot the data points after removing the fixed effects from the above regressions, we obtain Figure E-2. The red curves or lines are the quadratic or linear fit of the regression models. See the interpretation and discussion on these results in the introduction of the paper.

Table E-2: Panel Regressions of Level and Growth of R&D, Compustat 1962-2018

	(1) $\log RnD_{it}$	(2) $\Delta \log RnD_{it}$	(3) $\log RnD_{it}$	(4) $\Delta \log RnD_{it}$
$MarketShare_{it}$	0.100^{***}	-0.00240^{***}		
	(8.03)	(-4.11)		
$MarketShare_{it}^2$	-0.000762^{***}			
	(-6.83)			
$Duration_{it}$			0.0360^{***}	-0.00224^{***}
			(3.16)	(-3.18)
$Duration_{it}^2$			-0.000853^{***}	
			(-5.05)	
Firm f.e.	X		X	
Time f.e.	X		X	X
Industry f.e.				X
IndustryXyear f.e		X		
N	1,153	1,089	541	512
R^2	0.748	0.591	0.400	0.071

Note: This table shows the estimates of the regressions that relate the level and growth of R&D to market shares over the lifecycle of an eventual leader, as well as the estimates of the regressions that relate the level and growth of R&D to the age of leadership for leaders. Standard errors in parentheses are clustered by firm in model (1)-(3). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. For detailed description, see Appendix E.2.

Figure E-2: Share of R&D Conducted by Industry Leaders



Note: This figure illustrates the regression results. We plot the firm-year observations after removing all fixed effects from regression models (E-1) to (E-4) in Table E-2. Each panel corresponds to one regression. The top panels correspond to regressing the level and growth of R&D on market shares over the lifecycle of an eventual leader. The bottom panels correspond to regressing the level and growth of R&D on the age of leadership for leaders.