Symbiotic Competition and Intellectual Property

Rafael R. Guthmann and David M. Rahman*

February 22, 2023

Abstract

According to Nordhaus, the optimal life of a patent T^* trades off the "embarrassment" of monopoly with motivating innovation, given that imitators would otherwise copy inventions and, in competing with innovators, reduce their profit hence incentive to innovate. To test this argument, we develop an endogenous growth model with knowledge spillovers and imperfectly elastic labor supply. Innovators invent new varieties and earn monopolistically competitive rents during the life of a patent, but create deadweight losses in the labor market. When the patent expires, imitators copy innovators and competition ensues. Over time, firms continuously make follow-on "process" innovations and learn from each other, resulting in faster productivity growth relative to monopoly. Enough such innovations reverse the logic of patents: higher spillovers can imply a shorter T^* if, despite a diminution in product innovation, growth from process innovations matters more. Our calibration to the US economy offers T^* between 6 and 11 years, in contrast with the current global standard of 20. A counterfactual without spillovers yields $T^* \approx 34$, suggesting that lost opportunities for productivity growth from competition matter substantially for the optimal life of a patent. We also find that optimal patent policy depends crucially on the distribution of spillovers across industries. As such, we argue for a macroeconomic approach to patent policy instead of restricting attention to individual industry estimates.

---

*Universidad Alberto Hurtado and the University of Minnesota, respectively. We thank Illenin O. Kondo, Erzo Luttmer, Aldo Rustichini, Jan Werner, and the participants of the Minnesota Theory Workshop and the 44th meeting of the Brazilian Econometric Society for useful feedback.

"We believe in help for the underdog, but we want him to stay under."
Cry, The Beloved Country, Alan Paton, 1948.

1 Introduction

The centuries-old debate regarding the value of patents remains fascinating yet unsettled.[1] According to Nordhaus (1969), the optimal life of a patent T^* balances what Jefferson (1813) called the "embarrassment" of market power with incentives to innovate. Just what this trade-off looks like in fact is, of course, an empirical question. Some have argued that governments ought to eradicate patents for new inventions (Boldrin and Levine, 2013), while others suggest that patents' economic puzzle has not yet been solved (Moser, 2016; Williams, 2017). This paper contributes to the debate by evaluating patents in a general equilibrium framework that includes endogenous growth and dynamic competition with knowledge spillovers. Our main finding is that—while patents can, in principle, be socially harmful—they do, in fact, provide a valuable economic role. Our calibration to the US economy suggests a T^* between 6 and 11 years, in contrast with the current long-established global standard of 20. Very roughly, we find that high mark-ups[2] encourage firms to innovate even with short-lived patents.

Underlying Nordhaus's trade-off is the idea that a new product created by an innovator could be copied by an imitator at negligible cost, like a candle lighting another, until the new product market would experience enough competition that profits would be too low to warrant product innovation in the first place. In this argument, the imitator contributes no value beyond reducing market power; e.g., he does not lower the costs of production.

Our paper adds dynamism to the story: when an innovator invents a new product, an imitator copies its means of production while also making marginal, unpatentable "process" innovations continuously.[3] In turn, innovators build from imitators' inventions through "knowledge spillovers," leading to productivity growth driven by a virtuous cycle of mutual imitation and innovation. This cycle, which we dub "symbiotic competition," increases productivity and adds substantial opportunity costs to monopoly: in our calibration, $T^* \approx 34$ years without it.

[1] Instead of tersely summarizing the infinite literature on the history of patents and patent debates, old and new, we point to Moser (2013, 2016), Williams (2017), Bryan and Williams (2021) and references therein.

[2] See, e.g., De Loecker et al. (2018), De Loecker and Eeckhout (2018) and Haltiwanger et al. (2022).

[3] "There is no patent strong enough to cover ideas only *inspired* by the original invention, but the social value of the original invention surely ought to include that inspiration." (Bryan and Williams, 2021, p. 5.)

Symbiotic competition thus introduces a counterweight to the trade-off in the life of a patent. On the one hand, as argued fervently by Boldrin and Levine (2013), if it takes an imitator long enough to copy an idea, the window of time during which the innovator is a monopolist should be enough to motivate innovation, which they argue makes a compelling case against patents. Moreover, in line with Nordhaus's paradigm, the longer it takes to imitate, the less pressing is the need for patent protection. Contrariwise, symbiotic competition suggests a channel through which faster imitation may reduce the need for patents. Its "goose-gander" logic is this: if imitation is faster, productivity growth through mutual imitation and subsequent marginal innovations arrives sooner, which can increase welfare if the gain in productivity of existing varieties is large enough to compensate for the creation of fewer new varieties.

When spillovers vary across industries, so do the productivity gains from symbiotic competition. Therefore, absent industry- or sector-specific patent policy, or in case the spillover effects in a new product market are difficult to predict, T^* should depend on the entire distribution of spillovers, not just on those in any particular industry. Given the compelling evidence for such variation in spillovers,[4] we believe that this observation offers useful context for empirical work like Budish et al. (2015, 2016), which hones in on Nordhaus's trade-off and emphasizes industry-specific estimation of research elasticity with respect to patent duration. Not only is it important, we think, to understand this elasticity within a framework that includes mutual symbiotic spillovers, but also as part of a broad range of representative industries.

The optimal life of a patent varies markedly with spillover heterogeneity: $T^* \approx 24$ years when our spillover parameter takes its average value in every industry, but $T^* \approx 8$ when it varies across industries. This suggests that lost welfare from discouraging innovation in industries with high spillovers takes precedence over giving incentives to innovate in industries with low spillovers, as the latter does not contribute as much to welfare gains through lower costs of production. Thus, the importance of the argument above that higher spillovers can decrease T^* is amplified precisely by those industries with higher spillovers. To illustrate, consider Figure 1 below, which, using our calibration, compares welfare assuming constant spillovers

[4] Williams (2017) studies cumulative innovations and points to Galasso and Schankerman (2015) as well as Sampat and Williams (2019) for evidence of heterogeneity. Our calibration relies on Berlingieri et al. (2020), who estimate the distribution of spillovers across industries from about a dozen European countries.

across industries with heterogeneous spillovers. With constant spillovers, a patent of 20 years is close to maximizing welfare, whereas under spillover heterogeneity having no patents is only marginally worse than the status quo of 20 years. Regardless, a patent life of about 8 years is much better than both no patents and 20-year patents. (A similar result obtains with output, which is cardinal, instead of welfare, which is ordinal; see Figure 8.) This gives useful context to many patent policy debates, often divided between complete eradication and keeping the status quo, with pharmaceutical companies as poster children for the need for patents.

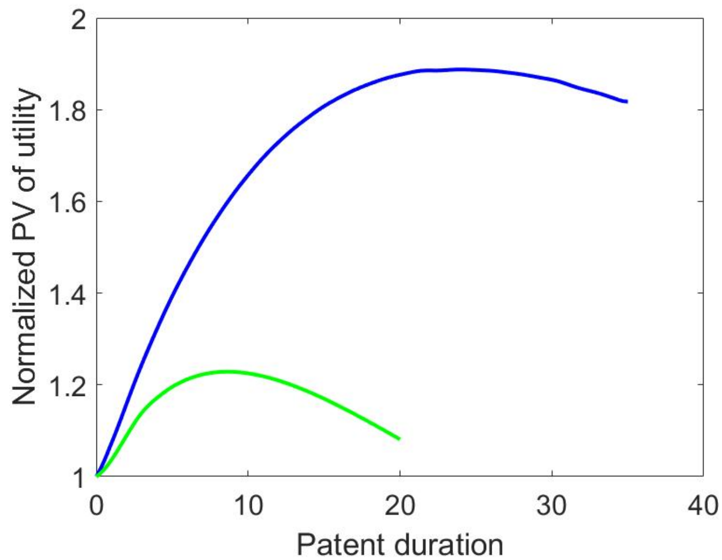

Figure 1: Welfare and patent policy with constant (blue) vs. heterogeneous (green) spillovers.

Symbiotic competition might also have a cumulative effect on aggregate economic growth: if start-up firms' initial productivity is proportional to the productivity of existing firms across industries then productivity growth within industries, compounded with cumulatively higher initial conditions for start-ups, creates a feedback loop that induces aggregate growth. Thus, endogenous growth may be fueled by economy-wide knowledge spillovers, as in Arrow (1962), Romer (1986) and Lucas (1988), further reducing the optimal life of a patent.

To summarize, in addition to Nordhaus' deadweight loss of monopoly, our calibrated model unveils three quantitatively significant determinants of the optimal life of a patent: within-industry spillovers, spillover heterogeneity across industries, and economy-wide knowledge spillovers. The incremental impact of each factor on T^* is shown in Figure 2 below.[5]

[5]The results in Figure 2 assume parameters that best fit a model with economy-wide and within-industry spillovers. Each individual factor is removed in turn without changing other parameters. See Section 4.

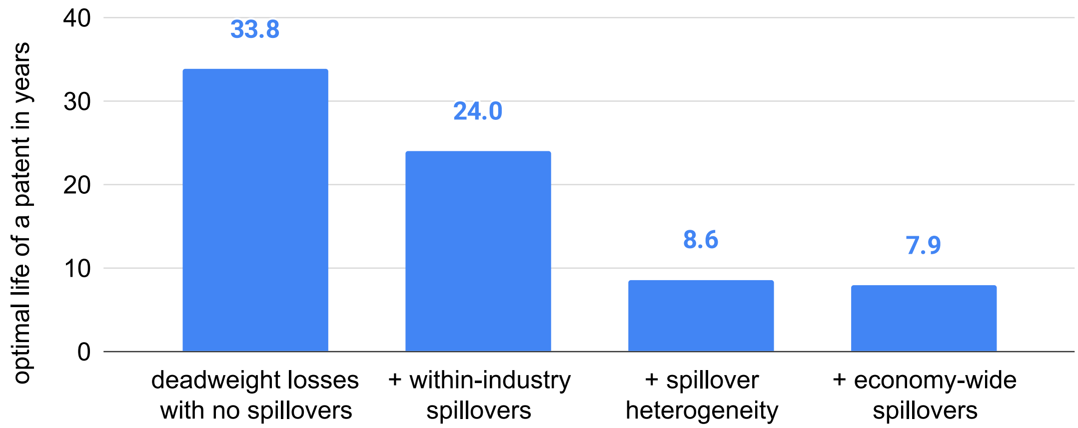

Figure 2: Incremental effects of spillovers on the optimal life of a patent.

Our main contribution in this paper is a new macroeconomic framework that brings together these three factors into a single quantitative model. To put our work into context, most endogenous growth models have innovators incorporate firms that compete monopolistically à la Dixit and Stiglitz (1977) with inelastic labor supply.[6] Despite rampant monopoly rents, no deadweight losses arise, as every monopolistic competitor sets the same markup, and thus factors are efficiently allocated across sectors. Therefore, if imitators are able to appropriate some of innovators' profits without adding value, they discourage innovation, which reduces welfare. For this reason, infinitely lived patents are typically optimal in these economies (e.g., Romer, 1986). To account for Nordhaus's trade-off between the deadweight loss of monopoly and innovation incentives, we assume that labor supply is imperfectly elastic, so market power causes inefficiently low wages, labor supply, and output. In our calibration, this labor-market distortion is enough to affect T^* significantly. Additionally, firms still make substantial profit, driven by high mark-ups, and households enjoy lower prices from cost-cutting innovations, driven by competition, when patents are short, as Figure 2 suggests.

In our model, once a patent expires, competition not only reduces deadweight losses via lower markups, but also increases productivity growth symbiotically, driven by laggard firms catching up to industry leaders. Abstractly, this idea is close to Liu et al. (2022), which builds on earlier work by Aghion et al. (2005), and Acemoglu and Akcigit (2012). However, there are important conceptual differences that distinguish our paper from these. The previous

[6] A recent example is Luttmer (2021), who studied the "commons" problem of duplication of research effort.

models endogenize process innovations by having firms choose how much to invest in them as a function of the productivity gap with their competitors. This yields an inverted-U relationship between competition and innovation. In our model, the starting point is a trade-off between making new products and improving existing ones. Although we could endogenize process innovations, it would be subsidiary to our main point that spillovers matter for patent policy. Our main insight—that productivity growth increases when firms are close together—is a natural feature of our environment. As such, process innovations inside an industry increase when firms are in neck-and-neck competition, whereas product innovations increase when neck-and-neck competition is postponed by patents. Overall innovations therefore exhibit an inverted-U relationship with competition for different reasons. Unlike the previous papers, which for tractability model innovations as jump processes, our paper has firms innovating continuously over time; the result of these innovations is sometimes positive but other times negative. Technically, we model firms' productivity rates as planar diffusion processes with rank-based characteristics, introduced by Fernholz (2002) and Fernholz et al. (2013) with mathematical finance applications in mind. The model is particularly tractable: transition densities for average productivity and the productivity gap lend themselves to analysis.

Of course, spillovers and "follow-on" innovations have been studied before, both theoretically and empirically.[7] Previous studies tended to focus on individual industries, like gene editing, to estimate how a patent may influence follow-on research. However, they did not explicitly model the social value of follow-on innovations, which drives our symbiotic effects. Our attempt in this paper is to look at the economy as a whole with a general equilibrium framework, so as to estimate the social costs of forgoing symbiotic competition during the life of a patent. In short, we estimate how follow-on innovations ought to influence patents.

The rest of the paper is organized as follows. Section 2 presents our model of within-industry symbiotic competition. Section 3 inserts it into an endogenous growth macroeconomic model. Section 4 reports on our calibration and estimation exercises. Section 5 discusses pertinent patent history and policy proposals, as well as the validity of competition in our model. Finally, Section 6 concludes. Proofs are relegated to the appendix.

[7] Again, see Scotchmer (1991); Bessen and Maskin (2007); Acemoglu and Akcigit (2012); Galasso and Schankerman (2015); Sampat and Williams (2019) and references therein.

2 Symbiotic Productivity Growth

In this section, we begin by describing our model of duopolistic productivity growth, which we interpret in terms of learning by doing with knowledge spillovers.

Consider an industry consisting of two firms, 1 and 2, competing in continuous time, indexed by $t \in [0, \infty)$. The (log-) productivity of firm i at time t is denoted by Z_{it}. This productivity parametrizes the firm's production function, to be discussed later. First, we present below a stochastic model of the evolution of Z_{it} for each firm i whose purpose is to tractably capture both learning by doing and knowledge spillovers.

Assumption 1. Each firm i's productivity is a stochastic process that satisfies

$$dZ_{it} = \begin{cases} (\mu + \theta)dt + \sigma dW_{it} & \text{if } Z_{it} < Z_{jt} \text{ and} \\ \mu dt + \sigma dW_{it} & \text{if } Z_{it} \geq Z_{jt}, \end{cases} \tag{1}$$

where j is the other firm in this industry and W_1, W_2 are independent Wiener processes.[8]

This model of productivity offers a simple way to understand the effect of spillovers on growth in a duopoly. The duopoly model of (1) above has the drift in a firm's productivity depending on whether it leads or lags the other firm. We interpret a leading firm's productivity as that of a monopolist's, with productivity growth due to learning by doing or research in process innovations captured by μ. At the same time, the productivity of a lagging firm grows at an even faster rate due to technological or other knowledge spillovers from the leading firm. This is captured by assuming that $\theta > 0$. Eventually, the laggard catches up and becomes the new leader, and any innovation by this leader is, in turn, absorbed by the new laggard. This yields overall productivity growth above and beyond the leader's rate.

The parameter θ captures how quickly firms catch up to their leaders when they lag. The limit of $\theta \to 0$ can be interpreted as there being no spillovers, where firms grow through innovations just as they would in the case of monopoly. On the other hand, as $\theta \to \infty$, the laggard firm's productivity converges towards the leader's arbitrarily quickly until, in the limit, each infinitesimal improvement is immediately absorbed by both firms.

[8] This two-dimensional stochastic process is characterized by having rank-dependent parameters—in this case, each dimension's drift. For more on rank-dependent diffusion processes, see, e.g., Fernholz et al. (2013).

Equation (1) generalizes the learning model of Lucas and Moll (2014) and its subsidiaries by acknowledging that technology absorption takes time. In such models, when two individuals meet, whoever has lower productivity immediately learns, or absorbs, the technology of the more productive individual and is henceforth able to produce at higher productivity. In our model, however, an individual's productivity can improve by, at most, θ per unit of time. In addition, unlike Lucas and Moll (2014), our model allows for the productivity of both leader and laggard to be subject to variation during their interaction. This simple extension generates rich economic dynamics. We interpret the parameter μ as a natural rate of learning by doing and σ as a parameter describing how a firm's productivity fluctuates randomly via process innovations. However, σ together with the catch-up parameter $\theta > 0$ opens the way for firms to outpace—and thus learn from—each other. We refer to the virtuous cycle that this induces as "symbiotic" productivity growth.

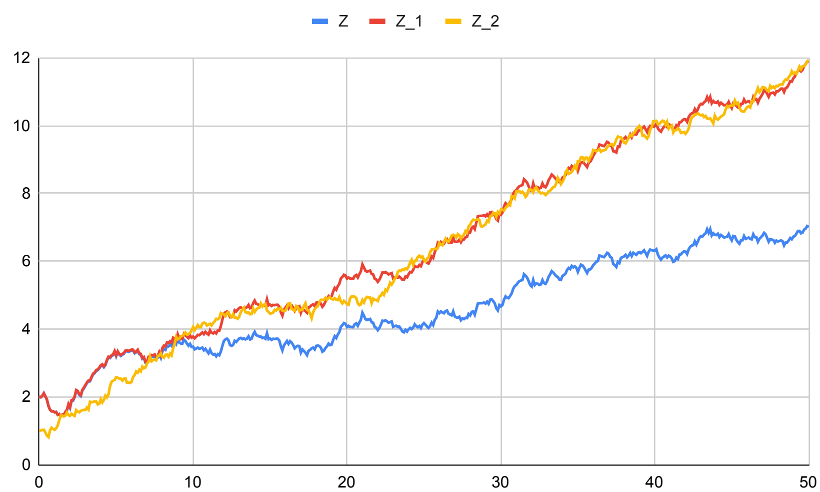

Figure 3: Sample path of productivity—monopoly versus duopoly.

To illustrate, Figure 3 above shows sample paths of productivity for each firm in the dynamic duopoly above, together with the path of a monopolist whose productivity Z has law of motion $dZ_t = \mu dt + \sigma dW_t$. The blue path shows productivity growing at rate $\mu = 0.1$ with $\sigma = 0.2$. The red and yellow paths reflect the system in (1) with $\theta = 0.25$. To facilitate comparison, the shocks to Z_1 are identical to those in Z.

Three immediate observations may be drawn from Figure 3. First, there is significant catch-up by the laggard firm to the leading firm. Secondly, productivity grows faster in duopoly than in monopoly. Thirdly, duopolists' productivity rates are usually close together, although they do sometimes veer away. Thus, even if firms usually compete neck-and-neck, they enjoy some periods of relative advantage over their competitors. These observations are formalized below by studying the transition density of each firm's productivity. The productivity process (Z_1, Z_2) above is naturally decomposed into sum and difference processes that submit to analysis more easily. Let us study them in turn.

Proposition 1. Average productivity $X = \frac{1}{2}(Z_1 + Z_2)$ *obeys the following law of motion:*

$$dX_t = (\mu + \tfrac{1}{2}\theta)dt + \sigma dW_{xt}$$

and $W_{xt} = \frac{1}{2}(W_{1t} + W_{2t})$ is a Wiener processes. Thus, average productivity in a duopoly grows at the rate $\mu + \frac{1}{2}\theta$, in contrast to the growth rate of a monopoly, μ.

Proposition 1 states that average productivity follows a relatively simple arithmetic Brownian motion. As such, its transition density into x at time t is given by

$$f_t(x) = \varphi_t\left(x - x_0 - (\mu + \tfrac{1}{2}\theta)t\right), \qquad \text{where} \qquad \varphi_t(z) = \tfrac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\tfrac{1}{2\sigma^2 t}z^2\right\}$$

is the PDF of a normal random variable with mean 0 and variance $\sigma^2 t$ and $x_0 = X_0$ is the initial value of X. Therefore, average productivity grows faster with knowledge spillovers than without them by $\frac{1}{2}\theta$. In this sense, knowledge spillovers increase productivity growth.

Proposition 2. *Firms'* productivity gap $Y = \frac{1}{2}(Z_1 - Z_2)$ *obeys the law of motion*

$$dY_t = -\tfrac{1}{2}\theta sgn(Y_t)dt + \sigma dW_{yt}$$

and $W_{yt} = \frac{1}{2}(W_{1t} - W_{2t})$ is a Wiener process independent of W_{xt}. Thus, the productivity gap between firms in a duopoly with knowledge spillovers drifts towards zero at rate $\frac{1}{2}\theta$.

The productivity gap between firms describes the extent of competition in an industry. By Proposition 2, this gap is a diffusion whose drift depends on its sign. If the difference in productivity is positive, it tends to diminish, whereas if it's negative, it will tend to increase. This is the "catch-up" effect due to spillovers, which leads firms to engage in "neck-and-neck" competition more frequently than they would without spillovers (i.e., if θ were equal to 0). As it turns out, Y has a transition density that can be written in closed form.

Lemma 1. *The productivity gap Y has transition density*

$$g_t(y) = \Psi_t(|y|) + \begin{cases} \varphi_t(y - |y_0| + \frac{1}{2}\theta t) & \text{if } y > 0, \text{ and} \\ e^{-\frac{\theta}{\sigma^2}|y_0|}\varphi_t(y - |y_0| - \frac{1}{2}\theta t) & \text{if } y \leq 0, \end{cases}$$

where φ_t is the normal PDF above, $y_0 = Y_0 \geq 0$ is the initial value of Y,

$$\Psi_t(|y|) = \frac{\theta}{2\sigma^2}e^{-\frac{\theta}{\sigma^2}|y|}\Phi_t(\tfrac{1}{2}\theta t - |y| - |y_0|), \quad \text{and} \quad \Phi_t(z) = \int_{-\infty}^{z} \varphi_t(x)dx$$

is the normal CDF of φ_t. For $y_0 < 0$, the transition density at y equals $g_t(-y)$.

Lemma 1 characterizes the transition dynamics of the productivity gap. The gap's probability density has two components: a term corresponding to the direct tendency from y_0 to y of an arithmetic Brownian motion with drift $-\frac{1}{2}\theta \operatorname{sgn}(Y)$, and a correction term for time spent hovering around zero due to catch-up (also known as Y's *local time* around 0) before escaping to y. To gain intuition for the dynamics of Y, it is useful to consider certain limits. As $y_0 \to 0$, the density g_t becomes symmetric, so it's just as likely that either firm is the leader or the laggard at every point in time. As $t \to \infty$, the industry's productivity gap has a stochastic steady state with a double exponential probability density g_∞, as the next result shows. This density describes the relative frequency of neck-and-neck competition in the long run, as well as the time firms spend ahead of their competitors, hence earning significant profit.

Proposition 3. *In the long run, firms' productivity gap has a probability density*

$$g_\infty(y) = \lim_{t \to \infty} g_t(y) = \tfrac{\theta}{2\sigma^2}e^{-\frac{\theta}{\sigma^2}|y|}.$$

Thus, the relative time during which firm 1's productivity lead exceeds $y \geq 0$ equals $\frac{1}{2}e^{-\frac{\theta}{\sigma^2}y}$. This duration converges to 0 at rate $\frac{\theta}{\sigma^2}$ as $y \to \infty$.

Another way to understand the predictions of this model is in terms of the distribution of catch-up time. That is, if y_0 is an industry's initial productivity gap between two firms, the *catch-up time* is the first time that Y hits zero: $\tau(y_0) = \inf\{t \geq 0 : Y_t = 0\}$. This is the first time the laggard firm catches up to the leader. Our next result characterizes its distribution.

Proposition 4. *For an initial productivity gap y_0, the catch-up time $\tau(y_0)$ has distribution*

$$\mathbb{P}[\tau(y_0) > t] = \Phi_t(|y_0| - \tfrac{1}{2}\theta t) - e^{\frac{\theta}{\sigma^2}|y_0|}\Phi_t(-|y_0| - \tfrac{1}{2}\theta t).$$

The catch-up time $\tau(y_0)$ distribution above corresponds to the first hitting time of an arithmetic Brownian motion. For intuition, recall that Liu et al. (2022), in a different model styled after Aghion et al. (2005), call the expected hitting time a measure of "market dynamism."

Together, average productivity X and the gap Y completely characterize firms' productivity, since $Z_1 = X + Y$ and $Z_2 = X - Y$. Since X and Y are independent by Proposition 2, the transition density of (Z_1, Z_2) equals the product of the densities of X and Y above.

Corollary 1. *The probability density function of (Z_1, Z_2) at time t is given by*

$$h_t(z_1, z_2) = f_t(x)g_t(y)$$

where $x = \frac{1}{2}(z_1 + z_2)$ and $y = \frac{1}{2}(z_1 - z_2)$.

Corollary 1 above delivers a closed form for the probability density function of firm productivity when laggards tend to catch up to leaders. Interestingly, even though the productivity gap is stationary, average productivity is not, so in the long run, there can be arbitrarily large differences in productivity across—but not within—industries.

3 Endogenous Growth with Patent Protection

In this section, we embed the previous model of symbiotic competition into an otherwise standard economy with endogenous "Romerian" growth, driven by a growing population's inventions of new product varieties. These varieties constitute inputs in the production of a final consumption good. Profits in such intermediate product markets are protected with patents. In each of these markets, there are either one or two firms producing, depending on the life of a patent for that product and the life of the variety itself.

3.1 Environment

We now describe the key ingredients of our model: households, intermediate and final good producers, blueprints for intermediate good varieties, and the patent system.

3.1.1 Households

Consider an economy populated by a continuum of identical households with mass $N_t > 0$ in continuous time $t \in [0, \infty)$, where $N_t = N_0 e^{gt}$ grows exogenously at rate $g > 0$. Households consume a "final" good and supply labor. They discount future consumption according to a common discount rate $r > 0$. A household's utility over output and labor streams is given by

$$\int_0^\infty e^{-rt} u_t(c_t, \ell_t) dt,$$

where c_t and ℓ_t stand for the household's level of consumption of the final good and labor at time t, respectively, and

$$u_t(c_t, \ell_t) = c_t - \tfrac{1}{2}\eta_t \ell_t^2$$

stands for the household's utility flow. The function u_t is linear in output c_t and quadratic in labor ℓ_t, so a household's marginal disutility of labor is linear. The coefficient $\eta_t > 0$ will be discussed later, when we come to the model's calibration. Labor can be directed towards (i) research and development of blueprints for new intermediate product varieties and (ii) the production of existing varieties of intermediate goods. We consider each in turn below.

3.1.2 Blueprints

Blueprints encode the ability to produce an intermediate good of some variety. They are created as follows. There is a continuum, with unit mass, of blueprint producing firms. A blueprint producing firm that employs $\ell_{rt} N_t$ units of labor for research during the interval of time $[t, t + dt)$, where $dt > 0$ is arbitrarily small, obtains a blueprint during this time interval with probability $\gamma \ell_{rt} N_t dt$. With probability $1 - \gamma \ell_{rt} N_t dt$, it does not. All such blueprint arrivals are iid across firms. The aggregate stock of blueprints evolves over time as a result of the creation of new blueprints and a natural rate of obsolescence of existing blueprints $\delta > 0$. Denoting this stock at time t by B_t, with $B_0 > 0$ given, it therefore obeys the law of motion[9]

$$\dot{B}_t = \gamma \ell_{rt} N_t - \delta B_t. \tag{2}$$

At each time t, we label the continuum of existing blueprints by $i \in [0, B_t]$.

[9] We rely on a heuristic law of large numbers to aggregate blueprints across a continuum of households. In the balanced growth path, blueprints will grow at the population growth rate g.

3.1.3 Patents and Intermediate Good Production

When a blueprint producer creates a blueprint at some time t_0, it acquires the technology to produce a new variety of intermediate good, call it $i \in [0, B_{t_0}]$. This firm then incorporates a new firm that produces a differentiated intermediate good of variety i, labeled $i1$, to reflect the fact that this firm is the first to produce it. At any time $t \geq t_0$, while the blueprint is not yet obsolete, firm $i1$ produces this intermediate good with the linear technology

$$y_{i1t} = A_{i1t}\ell_{i1t},$$

where y_{i1t} denotes output, A_{i1t} denotes the marginal product of labor, and ℓ_{i1t} denotes the amount of labor dedicated to the production of this intermediate good.

Blueprints become patented immediately upon creation. This implies that nobody else may produce the variety described by the blueprint while the patent is active. Patents last for a length of time $T \in [0, \infty]$. Therefore, while $0 \leq t - t_0 < T$ and the blueprint is not obsolete, firm $i1$ is the monopolistic producer of variety i. From time $t_0 + T$ on, however, firm $i1$ loses its monopoly grip: at time $t_0 + T$, an imitator firm, labeled $i2$, enters the market.

Unless blueprint i is obsolete, firm $i2$'s technology at any time $t \geq t_0$ is, similarly to firm $i1$, given by $y_{i2t} = \varepsilon_{it}A_{i2t}\ell_{i2t}$, where y_{i2t} is firm $i2$'s output, A_{i2t} is its marginal productivity of labor, ℓ_{i2t} its labor employed, and ε_{it} is an indicator of patent expiration:

$$\varepsilon_{it} = \begin{cases} 0 & \text{if blueprint } i \text{ is under patent at time } t\text{, and} \\ 1 & \text{otherwise.} \end{cases}$$

Thus, for each $j \in \{1, 2\}$, the ijth firm produces a differentiated intermediate good with one input, labor, and the linear technology above. Let $Z_{ijt} = \ln(A_{ijt})$. While the patent for blueprint i is active, Z_{i1} follows the law of motion

$$dZ_{i1t} = \mu dt + \sigma dW_{i1t}.$$

As soon as the patent expires, the market for variety i becomes a duopoly, and its firms' log-productivity profile (Z_{i1}, Z_{i2}) obeys the law of motion (1) of the previous section.

Assumption 2. $\theta \sim H$ for some distribution H with compact support.

Our calibration in Section 4 considers different possibilities for H, specifically the uniform and truncated versions of the lognormal and exponential distributions.

All of the stochastic processes above are assumed iid across varieties. Initial conditions for Z_{ij} are $(Z_{i1,t_0}, Z_{i2,t_0+T})$. We make two assumptions:

(1) When a new industry is set up, we assume that Z_{i1,t_0} is equal to a proportion $S \in [0,1]$ of the average Z of all industries of existing blueprints.

(2) Z_{i2,t_0+T} is such that the difference between Z_{i1} and Z_{i2} at $t_0 + T$ is distributed according to the long-run productivity gap, as described in Proposition 3. In our calibration, we also consider the situation where there is an expected productivity advantage of the incumbent relative to the entrant by subtracting the expected value of the advantage from the entrants' starting productivity (which, otherwise, would be distributed according to the long-run gap).

Their discussion is postponed until the calibration of the model.

3.1.4 Final Good Producers

There is a continuum, with unit mass, of producers of the final good to be ultimately consumed by households. At each time t, final good producers purchase intermediate inputs (y_{ijt}) and combine them to produce an amount y_t of the final good according to the CRS technology

$$y_t = \left[\int_0^{B_t} \left(y_{i1t}^{\beta} + \varepsilon_{it} y_{i2t}^{\beta} \right)^{\alpha/\beta} di \right]^{1/\alpha} \tag{3}$$

for some substitution parameters α and β. We assume that differentiated goods within a variety are at least as substitutable as goods across varieties: $\alpha \leq \beta < 1$. This production function reflects how differentiated intermediate goods produced by duopolists within the same variety may generally be imperfectly substitutable, yet also more sustitutable than differentiated goods of different varieties. The case $\beta = 1$ corresponds to the differentiated goods being perfectly substitutable. In this case, lower-cost firms charge the minimum of their competitors' cost and own monopoly price.

3.2 Equilibrium

In this subsection, we define general equilibrium in this economy. For the sake of clarity, we discuss goods and labor markets separately, but first, we offer the following general description.

Definition 1. An *equilibrium* consists of intermediate good prices and quantities (p_{ijt}, y_{ijt}),[10] wages (w_t), blueprints (B_t), output/consumption (y_t) and labor (ℓ_{rt}, ℓ_{pt}) such that:

1. Consumers maximize lifetime utility given wages (w_t) and dividends from firm profits.
2. Blueprint producers maximize profit by producing (B_t) with labor $\ell_{rt}N_t$.
3. Intermediate good producers set prices (p_{ijt}) to maximize lifetime profit given other firms' prices and final good producers' demand, using pricing strategies, when relevant, that constitute a Markov perfect equilibrium of the dynamic duopoly pricing game.
4. Final good producers maximize profit by producing (y_{ijt}, y_t) at prices (p_{ijt}).
5. All goods markets and the labor market clear.

Let us describe equilibrium more formally. First, consider consumers. Since their utility is linear in consumption, we may aggregate preferences and income into a representative household. The representative household's budget problem is given by

$$\max_{(c_t,\ell_t)\geq 0} \mathbb{E}\int_0^\infty e^{-rt}[c_t - \tfrac{1}{2}\eta_t\ell_t^2]dt \quad \text{s.t.} \quad c_t - w_t\ell_t \leq \Pi_t \quad \forall t,$$

where Π_t stands for dividends from the representative firm, defined below. For simplicity, we rule out any borrowing or lending, so—since goods cannot be stored—consumers face the budget constraint above at every time t. Optimizing behavior implies that per capita labor is given by $\ell_t = w_t/\eta_t$ and per capita consumption by $c_t = \Pi_t + w_t\ell_t = \Pi_t + w_t^2/\eta_t$.

To complete our description of equilibrium, we treat each of the above items separately next, beginning with final good producers' demand curves for intermediate goods.

[10] The final good will function as numeraire; without loss, its spot price is normalized to 1 at every time t.

3.2.1 Final Good Producers

Final good producers maximize the expected present value of profit by choosing inputs to purchase and output to sell at each time t, taking market prices as given and subject to constraint (3). In this model, final good producers' choices have no dynamic consequences, therefore maximization of lifetime profit is equivalent to maximization of flow profit at each time t. The next result takes final good producers' profit function and derives their demand curves for intermediate goods. Recall that the final good's spot price is normalized to 1.

Lemma 2. *Profit-maximizing final good producers demand each intermediate input y_{ijt} as a function of market prices and total quantity produced y_t according to*

$$y_{i1t} = y_t p_{i1t}^{-1/(1-\alpha)} \tag{4}$$

for a monopolist $i1$, that is, when $\varepsilon_{it} = 0$, and

$$y_{ijt} = y_t p_{ijt}^{-1/(1-\beta)} \left[p_{i1t}^{-\beta/(1-\beta)} + p_{i2t}^{-\beta/(1-\beta)} \right]^{-(\beta-\alpha)/[\beta(1-\alpha)]} \tag{5}$$

for a duopolist ij, that is, when $\varepsilon_{it} = 1$.

Thus, a monopolistically competitive intermediate good producer faces the isoelastic demand curve in (4) from own price. In a duopoly, demand for an intermediate good interacts with that for another in the same industry as in (5).

3.2.2 Equilibrium in the Intermediate Goods Market

Because labor productivity grows exogenously in this model, and firms are assumed to set prices in Markov perfect equilibrium, intermediate good producers maximize lifetime profit by maximizing flow profit period by period.[11] A firm $i1$ in an industry i still under patent chooses its price p_{i1t} to maximize flow profit, i.e., $\Pi_{i1t}^m = \max_p \{y_{i1t}(p)(p - c_{i1t})\}$, given wages w_t and aggregate output y_t, where $c_{i1t} = w_t / A_{i1t}$ is the marginal cost of production because $y_{i1t} = A_{i1t}\ell_{i1t}$ implies $w_t \ell_{i1t} = w_t y_{i1t} / A_{i1t}$, and $y_{i1t}(p) = y_t p^{-1/(1-\alpha)}$ by Lemma 2, and y_t is independent of firm $i1$'s choices.[12]

[11] The standard proof of this assertion is available on request.

[12] Since there is a continuum of intermediate goods, we assume that each one is too small to impact y_t.

Lemma 3. *Before its patent expires, producer $i1$ sets a monopoly price p_{i1t} satisfying*

$$\frac{p_{i1t} - c_{i1t}}{p_{i1t}} = 1 - \alpha.$$

Once a patent expires, duopolistic firms engage in Bertrand competition with differentiated commodities taking as given wages, demand, and aggregate output. Each firm ij's duopoly profit-maximization problem is given by $\Pi^d_{ijt} = \max_p\{y_{ijt}(p, q)(p - c_{ijt})\}$, where p stands for firm ij's price and q for firm ik's price, $k \neq j$. Equilibrium duopoly prices are given below.

Proposition 5. *After $i1$'s patent expires, equilibrium duopoly prices p_{i1t} and p_{i2t} satisfy*

$$\frac{p_{ijt} - c_{ijt}}{p_{ijt}} = \left[\rho_j(1-\alpha)^{-1} + (1-\rho_j)(1-\beta)^{-1}\right]^{-1} \qquad \textit{and} \qquad \rho_j = \frac{p_{ijt}^{-\frac{\beta}{1-\beta}}}{p_{i1t}^{-\frac{\beta}{1-\beta}} + p_{i2t}^{-\frac{\beta}{1-\beta}}}$$

for each firm $j \in \{1, 2\}$ in industry i at time t. A pricing equilibrium exists.

According to Proposition 5, a firm's price-cost markup is a weighted (harmonic) mean of intra-industry and inter-industry inverse elasticities of substitution, with weights ρ_j and $1 - \rho_j$. If $\beta = 1$, differentiated goods are perfect substitutes and duopoly prices are the minimum of the lowest monopoly price and the highest marginal cost.

Firm $i2$'s profits are paid to households as a dividend; the representative household gets

$$\Pi_t = \int_0^{B_t} \varepsilon_{it}\Pi^d_{i2t}di.$$

Firm $i1$'s profits are paid to blueprint producers who use them to pay for labor. This pins down wages, as we argue next.

3.2.3 Blueprint Producers and Labor Market Equilibrium

Let us clear the labor market. Labor is supplied by households who provide per capita labor ℓ_{pt} to firms for the production of intermediate goods as well as ℓ_{rt} for research into the development of new blueprints. On the demand side, intermediate good producers purchase labor at given wages in line with Lemma 3 and Proposition 5 above. At any time t, the amount of per capita labor demanded by the intermediate good producers is

$$L_{pt} = \int_0^{B_t} (y_{i1t}/A_{i1t}) + \varepsilon_{it}(y_{i2t}/A_{i2t})di.$$

In labor market equilibrium, $N_t \ell_{pt} = L_{pt}$. To find ℓ_{rt}, recall consumers' optimality condition:

$$w_t = \eta_t \ell_t = \eta_t(\ell_{pt} + \ell_{rt}).$$

Therefore, $\ell_{rt} = (w_t/\eta_t) - \ell_{pt} = (w_t/\eta_t) - L_{pt}/N_t$. To complete the model, we must reconcile the present value of profits from research with their labor cost. To this end, let $\bar{\Pi}_{i1t}$ be the expected present value of profit of a firm producing intermediate good $i1$ that was developed at time t. We assume that research firms are risk neutral, with the following profit function: $\gamma \ell_{rt} \bar{\Pi}_{it} - w_t \ell_{rt}$. Therefore,

$$w_t = \gamma \bar{\Pi}_{i1t},$$

where

$$\bar{\Pi}_{i1t} = \mathbb{E}_t \left[\int_t^{t+T} e^{-(r+\delta)(s-t)} \Pi^m_{i1s} ds + \int_{t+T}^{\infty} e^{-(r+\delta)(s-t)} \Pi^d_{i1s} ds \right]$$

and $\mathbb{E}_t$ denotes expectation given all information available at time t. Of course, $\bar{\Pi}_{i1t}$ itself depends on the wage schedule $(w_{s \geq t})$.

3.3 Balanced Growth Path

We restrict attention to *stationary equilibria*, where the distribution of intermediate good varieties by age is constant over time.[13]

Definition 2. A *balanced growth path* (BGP) is a stationary equilibrium where output grows at a constant rate.

In the balanced growth path, as the population grows at a constant rate, the labor supply and the stock of blueprints also grow at a constant rate, the distribution of firm ages is constant, and the growth rate of output per capita is constant as well. As blueprints depreciate at the rate δ, the balanced growth path requires a constant growth rate in the output of blueprints, so the stock of blueprints also grows at the same rate. Note that the population grows at a constant rate, and the production of blueprints is linear in labor. As feasible effort levels

[13] To obtain this equilibrium, at some initial period, we assume that the initial distribution of intermediate good varieties is the stationary one. This will necessarily include some arbitrarily old blueprints. This technical simplification is standard.

are in $[0, 1]$, to maintain a balanced growth path, the growth rate in the effort level must be zero; therefore, the balanced growth path is characterized by a constant effort level e^*, which implies that the stock of blueprints grows at the same rate as the population. Therefore, the growth rate in the balanced growth path is determined by the growth rate of population, g, and the distribution of growth rates in productivity of the industries for each specific variety. Therefore the stock of blueprints is of size $B(t) = b(t)/(\delta + g)$ and the cumulative distribution function $F(a) \in [0, 1]$ of blueprints by age in the balanced growth path is given by $F(a) = 1 - \exp(-(g + \delta)a)$.

The proposition below states that given the assumptions of our model, there exists a BGP:

Proposition 6. *There exists a balanced growth path.*

3.3.1 Patent policy

The policy lever that the planner in our model can use is constrained to a patent duration policy $T \in \mathbb{R}_+$. Thus, in our framework, the optimal policy is a patent length $T^* \geq 0$ that maximizes the utility of a household in the balanced growth path. Consider an economy in a balanced growth path at some date t. If $S > 0$, the growth rate of output depends on the patent policy as stated in the proposition below:

Proposition 7. *If $S = 0$, varying T does not change the growth rate in the balanced growth path. If $S > 0$, the growth rate in the balanced growth path is strictly decreasing in T.*

4 Quantitative Analysis

We begin this section by exploring the quantitative properties of our model. First, we describe our basic calibrated parameters. Then we report on the first simulations of our model under the simplifying assumptions of constant within-industry spillovers and none economy-wide. This model, being computationally more tractable, allows us to make detailed comparative statics exercises, which we report next. Afterwards, we discuss the results of calibrating our remaining parameters and final model estimation.

4.1 Basic Parameters

In order to prevent per capita labor supply from exploding in the balanced growth path (an individual cannot supply more than 168 hours per week), we consider a functional form for households' utility, $u_t(c, \ell)$, that implies a constant labor supply (so labor supply does not diverge to infinity or converge to zero). This utility function is specified as follows:

$$u_t(c, \ell) = c - B_t^{\frac{1-\alpha}{\alpha}} \ell^2.$$

Proposition 8. *Each household's labor supply is constant along a balanced growth path.*

Household disutility of labor depends on the stock of blueprints similarly to the way it depends on output of the final good. This assumption implies that the "leisure production" technology improves in proportion to the technology for producing goods. As a consumer might have more options for leisure when technology develops, this assumption means that technological progress increases the opportunity cost of labor in lockstep with labor productivity. Thus, technological progress does not change the supply of labor in equilibrium.

	Value	Meaning
		Externally calibrated parameters (rates are annualized)
g	0.95%	Growth rate of the population
r	3%	Discount rate
e	8%	Exit rate of firms
δ	4%	Depreciation rate of blueprints
μ	0.9%	Growth rate in productivity of a leading firm
θ	1.5μ	Expected productivity spillovers from leader to follower
		Internally calibrated parameters
σ	0.087	Size of the shocks to firms' productivity
β	0.74	Parameter for within-industry elasticity of substitution
α	0.60	Parameter for elasticity of substitution across industries

Table 1: Calibrated parameters given constant within-industry spillovers.

We externally calibrate six parameters and internally calibrate three, see Table 1. Consider each in turn. The population growth rate of $g = 0.95\%$ is consistent with US data over the past 3 decades. Firms' exit rate is set to $e = 8\%$, again in line with US data. For our initial calibration, we assume that θ is constant across industries and let $\theta = 1.5\mu$. Thus, the expected growth rate of productivity of laggard firms, $\mu + \theta = \mu + 1.5\mu = 2.5\mu$, is 2.5 times μ, the growth rate of leaders' productivity. This assumption is based on Berlingieri et al. (2020), an OECD study of the levels and growth rates in labor productivity for firms across OECD countries. Letting firms with above-median productivity be leaders and firms with below-median productivity be followers, Proposition 1 above implies that average productivity growth equals $\mu + \theta/2 = 1.75\mu$, and laggard growth equals $\mu + \theta = 2.5\mu$.

We assume a discount rate r of 3%, in line with typical estimates, which vary from 1.5% to 5%. In this subsection, we assume a depreciation rate of blueprints $\delta = 4\%$. Thus, technologies depreciate at a slightly lower rate than typically expected for capital equipment. In the next subsection, we use the size distribution of firms to identify the blueprint depreciation rate.

As reported in Table 1, we internally calibrate three parameters: (1) for the elasticity of substitution across varieties (α) and (2) between products from firms in the same industry (β), as well as (3) the shocks to firms' productivity σ, under the status quo policy (patent law grants 20 years of monopoly to inventors). We target the following parameters: (1) average growth rate of US GDP over the last 30 years: 2.5%, (2) average markup of around 50%, in line with various estimates of average US markups in recent decades, such as De Loecker et al. (2018), De Loecker and Eeckhout (2018), and Haltiwanger et al. (2022),[14] and (3) standard deviation of productivity across firms of 1/3, taken from Berlingieri et al. (2020). In the context of our model, the standard deviation of productivity across firms is the standard deviation of firms' productivity with respect to their industry's average, as the assumption of a linear production technology and homogeneous labor implies that average labor productivity is constant across industries. Table 2 compares model and target moments.

We choose $\mu = 0.9\%$, so the expected growth rate in productivity of individual intermediate

[14] There is substantial variation in average markups depending on the estimation method. While De Loecker et al. (2018) and De Loecker and Eeckhout (2018) suggest that recent average markups were very high, around 0.5 to 0.7 times marginal cost, other studies suggest actual markups are significantly lower, such as Vlokhoven (2022). Thus, we consider alternative calibrations targeting average markups between 0.30 and 0.70.

	Target	Model
Annual GDP growth rate	$\approx 2.5\%$	2.53%
Average markup	≈ 0.50	0.49
Standard deviation of productivity	33%	32.8%

Table 2: Model fit for targeted moments.

varieties is 1.575%, to match the aggregate growth rate of per capita income in the US. Otherwise, productivity growth for certain input varieties would be so slow that the relative prices of intermediate inputs would rise indefinitely relative to the final good.

4.2 Semi-endogenous Calibration

In this subsection we simulate our model under the simplifying assumption that $S = 0$, so new industries begin with log-productivity $Z_{1t} = 0$ that does not depend on the productivity of the other industries of the economy. In other words, we temporarily shut down economy-wide spillovers. We call this a "semi-endogenous" growth model (Jones, 1995) because, in this environment, growth in per capita output along the balanced growth path does not depend on patent policy. In the next subsection, we relax this assumption to estimate the model.

4.2.1 Model Results

Figure 5 below summarizes the outcome of our initial calibration with the parameters reported in Table 1 and $S = 0$. First of all, notice that blueprint production and average markups increase with patent length, but allocative efficiency decreases due to the "Nordhausian" deadweight losses that market power imposes on the labor market. This gives welfare, in terms of normalized household utility, an inverse-U relationship with patent length.

Notice also that a substantial production of novel technologies (a.k.a. blueprints) exists in equilibrium even when patents do not exist (that is, $T = 0$), and shifting from a policy of no patents to a policy where patents last 20 years increases the equilibrium production of blueprints by approximately 30%.[15] Thus, according to our model, patents promote technological innovation but are not strictly required for substantial innovative activity to prevail (as argued by various authors, such as Moser, 2013). Therefore, if monopolies had no social costs, the optimal patent law would be for patents to last indefinitely.

While patents increase innovation incentives, they also increase average markups and decrease average productivity by reducing the gains from symbiotic competition. There is a wide interval of patent policies such that both output and welfare increase from a regime of no patents ($T = 0$). Thus, our model refutes the claim by Boldrin and Levine (2013) that patents do not add value to society. However, in this calibration, the current policy of $T = 20$ years is suboptimal—an optimal policy consists of a shorter but positive patent duration.

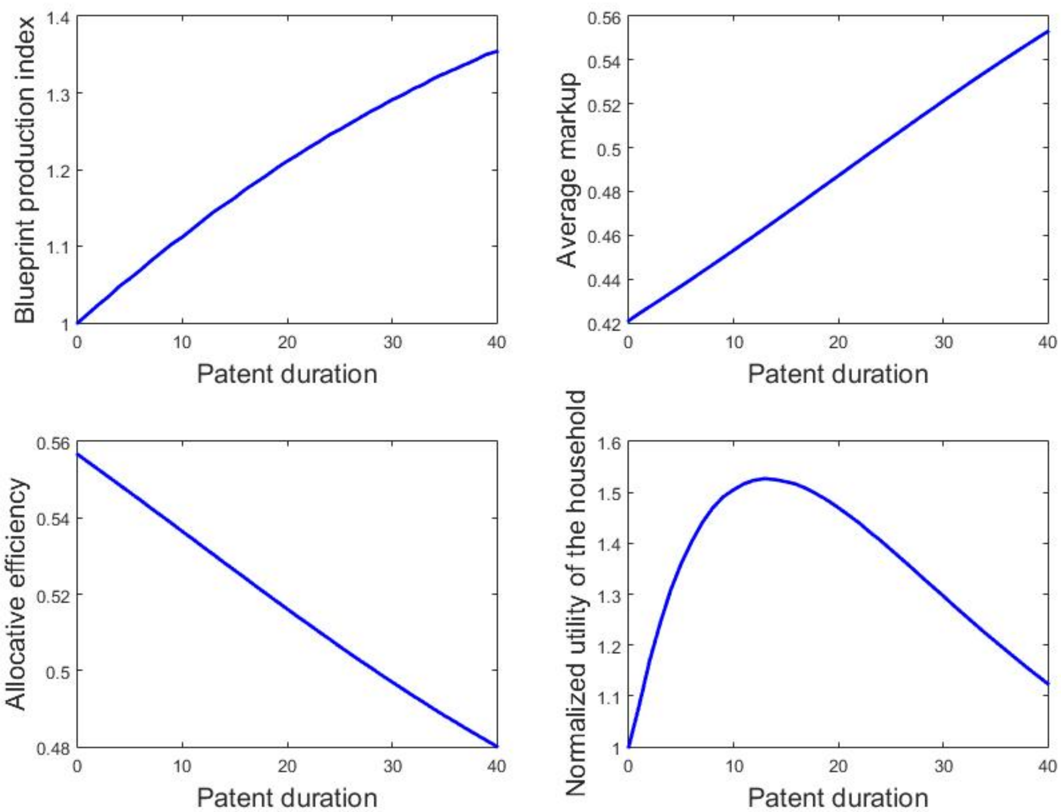

Figure 4: Longer patents increase blueprint production but decrease allocative efficiency, giving welfare an inverse-U relationship with patent duration.

[15]The blueprint index of Figure 5 normalizes to one the positive mass of blueprints when $T = 0$.

The welfare maximizing patent length is 13.4 years (which is approximately equal to the policy that maximizes the level of household consumption along the balanced growth path), and any patent duration in the interval from 11.0 to 16.7 years implies a level of household consumption within 1% from the maximum feasible level in the balanced growth path.

At the optimal patent length in this calibration, welfare is significantly higher than either of the two extreme policies of no patents ($T = 0$) and the status quo ($T = 20$). Moreover, the difference in welfare between $T = 0$ and $T = 20$ is small. This suggests that policy debates around the eradication of patents versus maintaining the status quo are difficult to reconcile because they lead to small differences in welfare, potentially subject to changes in the better policy (from the discrete set of $T = 0$ or 20) arising from small changes in parameters.

On the other hand, welfare increases by more than 40% from the status quo of $T = 20$ to the optimum of $T^* \approx 13$. Close to the optimum, welfare is stable: small changes in patent length around it do not affect welfare much. In addition, the optimal patent policy is robust with respect to small changes in parameters, and, around the optimal policy, welfare is robust, too.

Our calibration, with the parameters specified thus far, implies an average (instantaneous) price drop upon patent expiration of 34%. This is consistent with findings from empirical surveys of drug prices after the expiration of patents. Indeed, a meta-study by Vondeling et al. (2018) shows that, after only a year from patent expiration, drug prices can fall substantially: across three different studies, the measured fall is 17% to 32%, 34%, and 45%, respectively.

4.2.2 Comparative Statics

Let us consider how the optimal policy changes when we change the within-industry catch-up speed θ, the rate of time preference r, the population growth rate g, and the targeted markup rate. Our results show that the optimal patent duration varies substantially with different specifications of these parameters. However, T^* stays in the range of 8 and 20 years.

Figure 7 below reports these comparative statics. It shows that impatience lengthens optimal patent duration and population growth shortens it. Interestingly, there is enough variance σ in our calibration to conclude that as imitation becomes faster, i.e., θ/μ increases, optimal patent

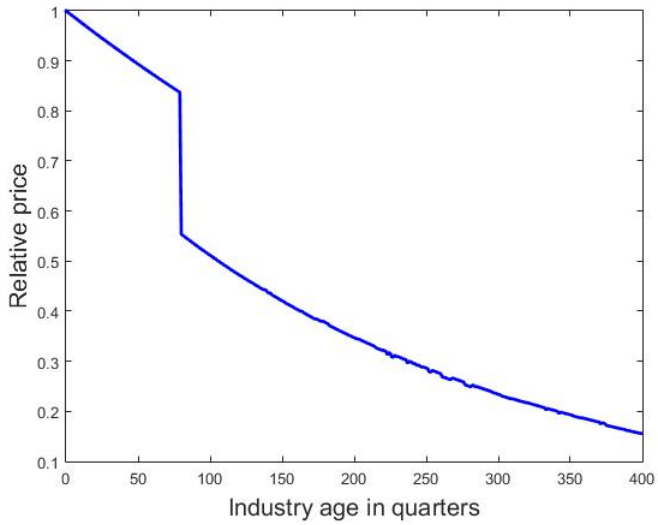

Figure 5: Example of expected prices posted across industry ages. Price drops discontinuously when a patent expires after 80 quarters.

duration decreases. This result runs counter the usual intuition behind patent protection that the threat of imitation justifies intellectual property. Intuitively, increasing θ increases the productivity gains from symbiotic competition, which increases total factor productivity and welfare compared to monopoly, hence, too, the opportunity cost of granting monopoly rights to new varieties. Is it reasonable to think that the optimal patent policy is to set $T = 0$? In our model, it is possible for the optimal policy to be that patents not exist at all if knowledge spillovers are sufficiently large. This provides a formalization of the argument in Boldrin and Levine (2013). For this to be the case, though, our calibration requires knowledge spillovers to be much higher than supported by the empirical evidence.

Finally, note that blueprint depreciation has a large, negative impact on patent policy. The negative slope is intuitive: if blueprints depreciate quickly then the future value of a patent is more heavily discounted, which reduces the benefit of patent protection. Because T^* is highly sensitive to the blueprint depreciation rate, it is important to carefully calibrate it.

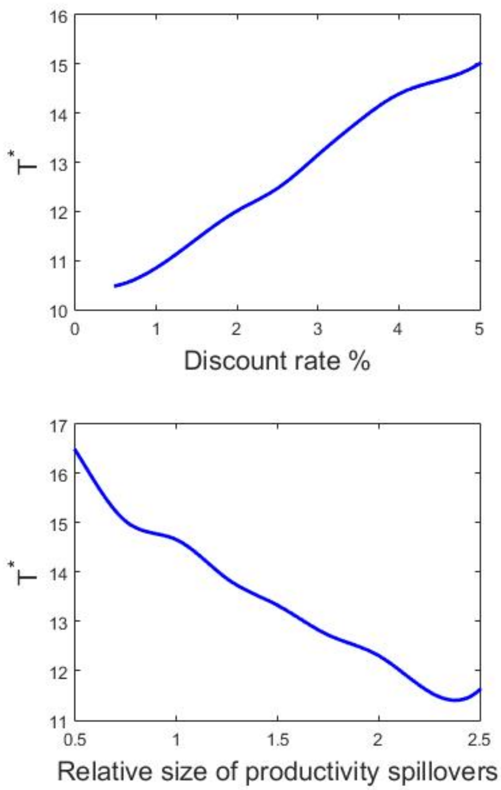

Figure 6: Effect on optimal patent policy of changes in the discount rate, the average targeted markup, the catch-up rate θ (keeping μ fixed), and the blueprint depreciation rate.

4.2.3 Spillover Heterogeneity

Next, we consider heterogeneous productivity spillovers across industries. We solve the model for spillovers distributed uniformly, exponentially and log-normally. In all cases, the support of the distribution is bounded below by 0, so there are no negative learning spillovers. Also, considering that we have not found empirical evidence that, in any industry, learning spillovers can be over ten times higher than the average productivity growth rate of leaders, we constrain θ to belong in an interval between 0 and 10 times leaders' productivity growth rate, μ.

The mean spillover rate $E[\theta]$ completely determines the uniform and exponential distributions (because a zero lower bound is imposed). In contrast, the log-normal distribution allows another degree of freedom via the variance. Fixing $E[\theta]$, we find that as varying the variance of the log-normal varies from the level of the uniform to the level of the exponential makes the optimal patent duration policy approximate first the optimal policy under the uniformly distributed θ and then under θ distributed according to the exponential distribution. Thus, the degree of variance of learning spillovers appears to be more important for our model's predicted optimal policy than the particular shape of the distribution of learning spillovers across industries. Table 3 presents our results from changing the distribution of θ.

Distribution of θ	Constant	Uniform	Log-normal	Log-normal(2)	Exponential
Relative variance of θ	0.00	1.00	1.01	2.98	3.00
Average markup	0.49	0.54	0.52	0.55	0.58
Price change at T^*	−34%	−34%	−34%	−35%	−33.9%
T^* in years	13.4	12.2	12.9	8.0	7.9

Table 3: Changes of model's output according to different distributions of learning spillovers

Two countervailing factors affect optimal policy when the variance of θ changes. On the one hand, increasing variance implies that more industries have high learning spillovers. In this case, some industries have much greater gains from symbiotic competition, which may lead to aggregate productivity gains if the gains from high-spillover firms exceed the losses from low-spillover industries due to the reduction in symbiotic growth from a longer patent. Therefore, this tends to increase the opportunity costs of patents, which in turn leads to a shorter optimal patent length.

On the other hand, however, a higher degree of learning spillovers also leads to a higher degree of intensity of competition and, therefore, lower profit margins once a patent expires. This increases the relative size of monopoly profits relative to duopolistic competition, which in turn makes patents more important to motivate innovation.

4.2.4 Incumbency Advantage

So far, in the quantitative analysis of the model, we assumed that there is no expected productivity gap between the entrant and the incumbent when the patent expires: the gap is distributed according to the long-run productivity gap, whose expected value is zero. Let us consider the following possibility: when the patent expires, entrants have an additional expected productivity gap with the incumbent. Our computational simulations show that T^* does not change very much: we consider increasing the expected log-productivity gap from 0 to 0.35 (so the entrant's expected productivity is ca. 57% of the incumbent's); this decreases T^* by approximately 0.7 years.

The reason why T^* decreases is clear: if incumbent firms have an initial advantage over prospective entrants, their profitability takes a strictly positive length of time to decrease and approximate its long-run level. Therefore, an incumbent's profit after a patent expires is slightly higher, which reduces the need for intellectual property protection. However, since the effect is relatively modest despite a very large initial productivity gap, we do not include this gap in our estimation of the model.

4.3 Model estimation

Following the quantitative exercises performed in the previous subsection, we now attempt a more precise estimation of optimal patent duration. We calibrate the model by estimating the degree of heterogeneity in productivity spillovers and the magnitude of economy-wide productivity spillovers, as well as incorporating in our identification strategy the model's predictions of the distribution of firm sizes.

When $S \in (0, 1]$, long-run economic growth does not require population growth: even with zero population growth, the balanced growth path exhibits positive growth in output per capita, driven by the accumulation of total factor productivity gains. We call this situation one of "endogenous TFP growth." All else equal, average productivity and output grow over time at a faster rate than when $S = 0$ and initial log productivity $Z_{1t} = 0$, as assumed in Section 4.2, because firms' initial conditions now gradually improve.

Our model estimation consists of finding parameters (Table 4) that best match four targeted moments (Table 5) together with the size distribution of firms (Figure 8). To this end, we re-calibrate the model separately for each S in a grid of values, assuming that θ is log-normally distributed across firms. We chose the log-normal distribution because it gives us an additional degree of freedom in terms of the variance of θ across industries, compared to the exponential distribution. Moreover, because of our zero lower bound on θ (we don't allow firms to have "negative" learning spillovers), with a uniform distribution the mean and variance are inextricably linked, limiting degrees of freedom. Table 4 below presents the model's calibrated parameters under the assumption that spillovers are log-normal.

	Value	Meaning
		Externally calibrated parameters (rates are annualized)
g	0.95%	Population growth rate
r	3%	Discount rate
e	8%	Exit rate of firms
μ	0.9%	Growth rate in productivity of a leading firm
$E[\theta]$	1.5μ	Productivity spillovers from leader to follower
$[\underline{\theta}, \overline{\theta}]$	$[0, 10\mu]$	Support of distribution of productivity spillovers
		Internally calibrated parameters
σ	0.068	Size of the shocks to firms' productivity
β	0.69	Parameter for within-industry elasticity of substitution
α	0.63	Parameter for elasticity of substitution across industries
δ	2.5%	Depreciation rate of blueprints
σ_θ^2	0.45	Variance parameter of the log-normal distribution of θ
S	0.07	Aggregate productivity spillovers to the new industries

Table 4: Model calibration for the balanced growth path under status quo policy.

In line with our previous calibration, we target the following four moments: GDP growth, average markups, the variance of productivity and the price drop upon patent expiration (see Figure 6). Table 5 below reports the target moments and the calibrated model's results.

In addition, we look for parameters that best approximate the size distribution of firms. These include the variance of knowledge spillovers (σ),[16] the depreciation rate of blueprints (δ),[17]

[16] An increase in the variance of spillovers increases the variance of productivity growth rates, hence, too, the variance of firm sizes.

[17] The lower the depreciation rate, the higher the proportion of old industries that have a longer time to grow, and thus, the higher the proportion of large firms in the economy.

	Target	Model
Annual GDP growth rate	$\approx 2.5\%$	2.50%
Average markup	≈ 0.50	0.477
Standard deviation of productivity	33%	32.0%
Price drop upon patent expiration	$\approx$ 25-35%	32.4%

Table 5: Model fit for targeted moments.

and the productivity parameter S of startup industries.[18] As we dsicussed previously, S also influences the growth rate of output, though, which we target to be approximately 2.5%, while the variance of knowledge spillovers influences the standard deviation of productivity across firms (targeted at 33%) as well as the average markup of the economy (target at 50%). The depreciation rate of blueprints does not affect the growth rate of the economy, since it only determines the distribution of industry ages in the balanced growth path. It also does not greatly affect the distribution of productivity gaps or average markups.

In our calibration, we target the distribution of firm sizes in terms of employment in the US in Axtell (2001): the number of firms larger than 5, 10, 25, 50, 100, 300, 1,000, 2,500, 5,000, and 10,000 employees given the total number of firms in the population. In our calibration, we compute the CCDF (1 minus the CDF) of firm sizes generated by the model and measure the fit of this distribution with respect to the CCDF implied by the data. To measure the fit of the model, we normalize the number of employees of a firm to be the unit of the quantity of labor employed by a firm that minimizes the average square of the log deviation of the CCDF of firm sizes between the model and the data. Then we choose the model's parameters to minimize this residual deviation. Our calibrated model attains an average log square deviation from the data of 10%. The target and model size distributions are shown in Figure 8 below.

4.4 Calibration Results

Our calibrated model suggests an optimal patent life of 7.9 years. As displayed in Figure 1 at the beginning of the paper, our calibrated model suggests that a policy of no patents provides

[18] The higher S is, the larger the size of firms in new industries relative to old industries will be.

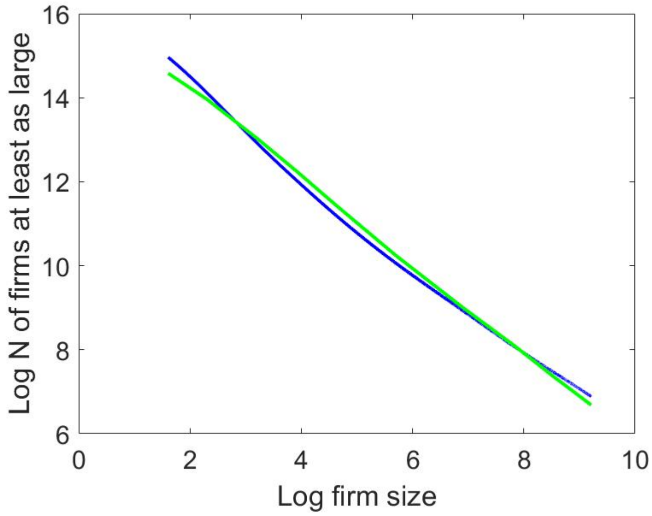

Figure 7: Distribution of firm sizes according to employment, data (blue) vs. model (green).

households with lower welfare than the status quo of 20 years. This result contradicts the claim by Boldrin and Levine (2013) that intellectual property law, as currently practiced, has a net social cost relative to the policy of abolishing patents.

If we do not consider the functional form of the household's utility function and just seek to maximize the present value of consumption along the balanced growth path, then Figure 8 shows that a patent duration policy of 8.0 years yields maximum consumption and that any patent duration policy between 6.4 and 11.2 years leads to household consumption within 1% of the level attained under the optimal policy. The consumption-maximizing T is only slightly different from the utility-maximizing policy, suggesting that our model's policy prescriptions do not heavily depend on the functional form of the household's utility function. Compared to the policy of abolishing patents, consumption is 28% higher in the balanced growth path at the optimal policy of 8.0 years and is 12% higher at the status quo policy of 20 years.

The decomposition of different effects on $T^* = 7.9$ is presented in Figure 2 in the introduction. Fixing all remaining parameters to the value of this calibration, we measure the optimal policy excluding aggregate productivity spillovers (that is $S = 0$), additionally excluding heterogeneity in θ (i.e., $\theta = 1.5\mu$), and finally, we estimate T^* under the assumption that symbiotic competition does not exist (i.e., $\theta \equiv 0$), which implies that the only cost of patents

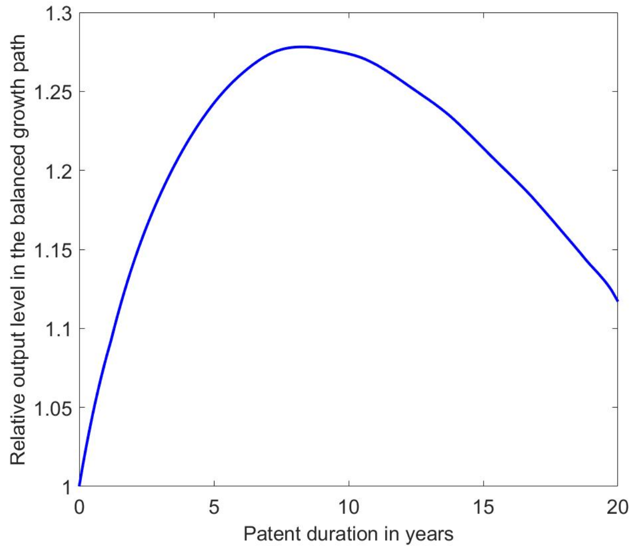

Figure 8: The effects of patents on final consumption.

is due to deadweight losses from monopoly power in the labor market.[19]

5 Discussion

In this section, we discuss some policy background, possible extensions, and existing debates. We also discuss the possibility of endogenous industrial organization by considering a monopolist who could mimic symbiotic growth with two divisions in "pseudo-competition."

5.1 Current and Possible Patent Policies

Current patent policy is governed by the "Agreement on Trade-Related Aspects of Intellectual Property Rights," or TRIPS, reached towards the end of the Uruguay Rounds around 1990 by all members of the WTO. There are many important facets to patent policy, such as interpretation, scope, and enforceability, and the literature on patent policy is, of course, boundless. Our analysis concentrates on a specific but fundamental aspect of patents: their

[19]Since the distribution of productivity gaps is undefined if $\theta \equiv 0$, we assume that the productivity gap is distributed in the same way as in the fully calibrated model, which yields the same degree of market power.

duration.[20] Although the life of a patent has evolved over the centuries since TRIPS, it was negotiated to be 20 years. TRIPS also instituted a non-discrimination policy that granted equal protection across industries. Given the protracted nature and multinational character of negotiations, it is not at all clear what ultimately led to the decision of non-discrimination and 20-year patents. Certainly, it is not clear a priori whether this number is close to being optimal, what criteria it fulfilled, or even the criteria it was intended to fulfill.

Policy debates and proposals often center around the recovery of research investments. For instance, according to Becker (2013), *[e]ven pharmaceutical and biotech companies, the main examples where patents are clearly necessary to encourage innovation, usually do not need more than about a decade of monopoly power to encourage their very large investments in new drugs. This is the case in many actual examples where after about ten years, molecularly similar drugs often are patented and compete against drugs with the original patents.* We view Becker's opinion as broadly aligned with our model, seemingly arguing that the *effective* life of a patent is about ten years, and this is plenty to motivate innovation. However, Becker's argument does not explicitly include the social benefits of imitation, such as lower costs of "molecularly similar drugs." Our model does, and, importantly, does so quantitatively.

It has been argued elsewhere that optimal patent policy ought to be discriminatory, that is, differ across industries.[21] For instance, the huge sunk costs of pharmaceuticals comes up repeatedly in analyses of optimal patents to emphasize the need for patent policies that differ across industries. Indeed, Posner (2012) writes, *[w]ith some exceptions, U.S. patent law does not discriminate among types of inventions or particular industries. This is, or should be, the most controversial feature of that law. The reason is that the need for patent protection in order to provide incentives for innovation varies greatly across industries.* We agree with Posner on this incontrovertible observation. Arguably, for this reason, there are stark differences between copyright and patent law. Although, ideally, patent rules would be tailored to each individual industry, there are compelling reasons for some uniformity in

[20] For a fascinating summary of the history of patent duration in the US prior to TRIPS, as well as the negotiations that led to TRIPS, see Lester and Zhu (2018).

[21] For a sample of economic discussions, see, e.g., Scherer (1972), Klemperer (1990) and, more recently, Acemoglu and Akcigit (2012). For a legal perspective, see, e.g., Carroll (2005).

patent law. Pragmatically, negotiating patent law worldwide can be incredibly costly and multifaceted (Stewart, 1993). Economically, some aspects of innovations, such as the extent of spillover effects in an industry, may not be known or knowable a priori. These issues point toward uniform patents.

Of course, a more sophisticated patent policy, able to adapt to different industries, could in principle improve welfare, and it would be possible to study optimum discriminatory patent policy within the framework presented in this paper. For example, we might consider patents that differ according to θ, that are time-dependent, or that track actual profit. One might also consider expanding the set of instruments available to motivate innovation, such as research subsidies. Our main contribution in this paper, as we see it, is to begin a quantitative macroeconomic study of patents that incorporates knowledge spillovers and growth together with the significant deadweight losses that monopoly incurs. As such, we believe that studying non-discriminatory patent policy is a good place to start, and leave an analysis of more sophisticated policies of intellectual property for the future.[22]

Why not patent process innovations, too? According to Posner (2012), *[i]n an industry in which teams of engineers are employed on a salaried basis to conduct research on and development of product improvements, the cost of a specific improvement may be small, and when that is true it is difficult to make a case for granting a patent.* Although we agree with Posner on this point, it is also true that a large amount of patents are issued for process innovations. Our argument, of course, is not that we should do away with all process innovations. As Posner claims, many innovations cannot reasonably be patented, but the accumulation of these small innovations can add up to a lot. Our model captures and emphasizes these small innovations and imitations as an integral part of the problem with patents.

[22]One assumption that vindicates the use of non-discriminatory policies is uncertainty over the extent of knowledge spillovers. On the one hand, it is not clear whether knowledge spillovers are known a priori in a new industry. It may be the case that an initial product innovation is easy to copy but subsequent process innovations are not. Knowledge spillovers are likely to differ substantially across industries and are potentially endogenous due to investment from both innovators and imitators. Our approach simplifies this aspect of the problem by finding the optimal life of a patent when θ is unknown ex ante. That is, our non-discriminatory patents are assumed to motivate researchers who draw θ after they invent a blueprint.

It has also been argued that the patent system is weaponized to the detriment of society. As pointed out by Boldrin and Levine (2013), "patent trolling" is often undertaken by "dying firms" to exploit market power and reduce follow-on innovation. Our model computes the optimal *effective* life of a patent that grants monopoly power to an innovator. This power may, of course, include patent trolling, which should factor into the *statutory* life of a patent. Again, we postpone a formal analysis of patent trolling for the future.

Finally, there has been a lot of debate over decades about the institution of patents itself, that is, whether to abolish it outright. Our analysis provides a more nuanced view. Rather than argue against patents on a moral level, the economic argument appears to be as follows according to our Figure 1. In our calibration with constant spillovers, a 20-year patent is a better outcome for society than no patents at all, while the optimum life of a patent is around 15 years. With calibrated spillover heterogeneity, no patents are superior to 20-year patents, but the optimal life of a patent is still strictly positive at around 5 years.

5.2 Endogenous Industrial Organization

We have assumed that when there are two firms in an industry, each firm's productivity is subject to independent stochastic shocks, and if one firm is lagging behind, it can learn faster than the productivity leader. These two assumptions imply that firms' expected average productivity in the duopolized industry grows at rate $\mu + \theta/2$, but if the industry is monopolized, expected productivity grows only at rate μ, which is of course strictly lower.

Profits would increase if production costs decreased due to improved productivity. Thus, it would be rational for a monopolist to emulate duopolistic technological progress if it could increase productivity growth. In this subsection, we consider the case where a monopolist can do so by hiring a manager to run part of his company as a separate division. We interpret this firm as running two divisions in "pseudo-competition," where one unit learns from the other as if they were a duopoly. The monopolist uses the leading technology of the two divisions to produce output and replicate the technology of symbiotic competition in our environment; he commits to employing the additional division forever after starting it.

Now, the monopolist's expected productivity growth is bounded above by that of an industry under duopoly: expected productivity grows by at most $\mu + \frac{1}{2}\theta$.[23] Moreover, the gain in expected profit from having two divisions instead of one is the gain in productivity over time. Let $\Pi^{m,t}_{jt+\Delta}$ be the flow profit at time $t + \Delta$ if the monopolist has two divisions instead of one, with the second division having been created at time $t \leq t + \Delta$. Clearly, if $\Delta = 0$ then $\Pi^{m,t}_{jt+\Delta} = \Pi^{m,t}_{jt}$ equals Π^{m}_{jt}, the one-division monopoly profit at time t, since productivity gains from symbiotic competition between the two divisions take time to accrue. If t is the time when the second division is created, expected profit at time Δ after t is at most

$$\mathrm{E}_t[\Pi^{m,t}_{jt+\Delta}] = y_t(w_t/\alpha)^{-\lambda}(1-\alpha)\exp(\lambda z_{jt} + \Delta(\mu + \theta/2)),$$

where z_{jt} is the log productivity of the monopolist's more productive division. On the other hand, the expected profit of a monopolist without an additional division equals

$$\mathrm{E}_t[\Pi^{m}_{jt+\Delta}] = y_t(w_t/\alpha)^{-\lambda}(1-\alpha)\exp(\lambda z_{jt} + \Delta\mu).$$

Let $G_t(\Delta)$ be the monopolist's maximum expected profit gain after a length of time Δ from starting the additional division at time t:

$$\begin{aligned}
G_t(\Delta) &= \mathrm{E}_t[\Pi^{m,t}_{jt+\Delta}] - \mathrm{E}_t[\Pi^{m}_{jt+\Delta}] \\
&= y_t(w_t/\alpha)^{-\lambda}(1-\alpha)[\exp(\lambda z_{jt} + \Delta(\mu + \theta/2)) - \exp(\lambda z_{jt} + \Delta\mu)] \\
&= y_t(w_t/\alpha)^{-\lambda}(1-\alpha)\exp(\lambda z_{jt} + \Delta\mu)[\exp(\Delta\theta/2) - 1].
\end{aligned}$$

Assume, however, that the monopolist cannot set up this additional division for free: he has to hire a manager to run this division. Let C_{jt} be the flow cost the monopolist has to pay to the manager for him to manage his additional division in the industry for variety j at period t. We interpret this C_{jt} as representing the opportunity cost of keeping the manager from exiting the firm and setting up his own firm in the industry, competing with the former monopolist. Under this interpretation, the monopolist has all the bargaining power and makes a take-it-or-leave-it offer to the manager. As such, the manager accepts managing the division for a value C_{jt} if and only if C_{jt} is greater than or equal to expected flow profits of one firm under duopoly, assuming, of course, that the technology of the two firms in the industry under

[23] A monopolist with two divisions of unequal productivity will have its leading division grow at rate μ until the lagging division catches up to the leading one. From then on, expected growth equals $\mu + \theta/2$.

duopoly is the same as the technology the monopolist has with two divisions. If the firm is under constant threat of renegotiation with the manager, long-term contracts fail. As such, the firm must trade-off the long terms benefits of growth from two divisions with the short term costs of compensating division managers due to their ability to quit and compete in the industry. A sufficiently patient firm, therefore, can mimic symbiotic productivity growth due to competition, but not an impatient one, as the next result shows.

Proposition 9. *There exists a discount rate* $\underline{r} > 0$ *such that for all discount rates higher than* $\underline{r}$*, the monopolist does not replicate symbiotic competition in equilibrium.*

Intuitively, this result implies that competition allows the benefits of patience to be replicated by myopic agents: patient firms have natural incentives to value the future benefits of faster technological progress. For inpatient firms, long-term growth is heavily discounted, so they do not take it into consideration. Competition, however, induces the faster technological growth that occurs under a inpatient monopolist, even if firms are completely myopic.

6 Conclusion

This paper develops a theoretical framework that interacts competition and economic growth both from the extensive margin of product innovation and the intensive margin of process innovation, or productivity improvements, to analyze key trade-offs for patent policy. Patents provide an incentive for innovation by securing monopoly rights for new products; our analysis allows us to estimate the costs of these monopolies along two dimensions: (1) by suppressing competition, monopolies in new products slow down productivity improvements of existing technologies via follow-on process innovations, and (2) monopoly power induces the usual static deadweight losses. In addition, we identify the importance of spillover heterogeneity in evaluating optimal policies that are constrained to apply uniformly across industries.

Our calibration shows that learning spillovers are a major factor in driving economic growth: under neck-and-neck competition, learning spillovers drive the productivity frontier and enable faster growth than in monopolized industries. Therefore, competitive markets add value not only by providing allocative efficiency in a static setting but also by producing a sustained increase in total factor productivity growth. Thus, patents create incentives for innovation but preclude the gains of symbiotic competition from materializing.

Our results run counter to the claim that shifting from our current patent system to a policy of no patents would improve welfare. Our calibration suggests an optimal patent length between 6 and 11 years. Further structural empirical work is needed to obtain detailed estimates of the optimal life of a patent. Nevertheless, our goal in this paper was to argue that a quantitative macroeconomic perspective is not just possible but critical for the study of patents.

References

Daron Acemoglu and Ufuk Akcigit. Intellectual property rights policy, competition and innovation. *Journal of the European Economic Association*, 10(1):1–42, 2012. 4, 5, 32

Philippe Aghion, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728, 2005. 4, 10

Kenneth J. Arrow. The economic implications of learning by doing. *The Review of Economic Studies*, 29:155–173, 1962. 3

Robert L. Axtell. Zipf distribution of u.s. firm sizes. *Science*, 293(5536):1818–1820, 2001. 29

Gary Becker. On reforming the patent system—becker. `https://www.becker-posner-blog.com/2013/07/on-reforming-the-patent-system-becker.html`, 2013. The Becker-Posner Blog, 07/21/2013. 32

Giuseppe Berlingieri, Sara Calligaris, Chiara Criscuolo, and Rudy Verlhac. Laggard firms, technology diffusion and its structural and policy determinants. 2020. OECD. 2, 20

James Bessen and Eric Maskin. Sequential innovation, patents, and imitation. institute of advanced study school of social science working paper 25. 2007. forthcoming. *RAND Journal of Economics*, 2007. 5

Michele Boldrin and David K. Levine. The case against patents. *Journal of Economic Perspectives*, 27:3–22, 2013. 1, 2, 22, 24, 30, 34

Kevin A. Bryan and Heidi L. Williams. Innovation: Market failures and public policies. In Kate Ho, Ali Hortacsu, and Alessandro Lizzeri, editors, *Handbook of Industrial Organization 5(1)*, pages 281–388. Elsevier, Amsterdam, Netherlands, 2021. 1

Eric Budish, Benjamin N Roin, and Heidi Williams. Do firms underinvest in long-term research? evidence from cancer clinical trials. *American Economic Review*, 105(7):2044–85, 2015. 2

Eric Budish, Benjamin N Roin, and Heidi Williams. Patents and research investments: Assessing the empirical evidence. *American Economic Review*, 106(5):183–87, 2016. 2

Michael W Carroll. One for all: The problem of uniformity cost in intellectual property law. *Am. UL Rev.*, 55:845, 2005. 32

J. De Loecker, J. Eeckhout, and Unger G. The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 2018. 1, 20

Jan De Loecker and Jan Eeckhout. Global market power. Working paper, 2018. 1, 20

Avinash K Dixit and Joseph E Stiglitz. Monopolistic competition and optimum product diversity. *The American economic review*, 67(3):297–308, 1977. 4

E Robert Fernholz. Stochastic portfolio theory. In *Stochastic portfolio theory*, pages 1–24. Springer, 2002. 5

E. Robert Fernholz, Tomoyuki Ichiba, Ioannis Karatzas, and Prokaj Vilmos. Planar diffusions with rank-based characteristics and perturbed tanaka equations. *Probability Theory and Related Fields*, 156:343–374, 2013. 5, 6, 41

Alberto Galasso and Mark Schankerman. Patents and cumulative innovation: Causal evidence from the courts. *The Quarterly Journal of Economics*, 130(1):317–369, 2015. 2, 5

John C Haltiwanger, Julia S. Foster, and Cody Tuttle. Knowledge growth and the allocation of time. *NBER working paper 30491*, pages 1–64, 2022. 1, 20

J. Michael Harrison. *Brownian Motion and Stochastic Flow Systems*. John Wiley and Sons, New York, NY, United States, 1985. 42

Thomas Jefferson. Letter to Isaac McPherson, August 13. 1813. 1

Charles I Jones. R & d-based models of economic growth. *Journal of political Economy*, 103 (4):759–784, 1995. 21

Paul Klemperer. How broad should the scope of patent protection be? *The RAND Journal of Economics*, pages 113–130, 1990. 32

Simon Lester and Huan Zhu. Rethinking the length of patent terms. *Am. U. Int'l L. Rev.*, 34:787, 2018. 32

Ernest Liu, Atif Mian, and Amir Sufi. Low interest rates, market power, and productivity growth. *Econometrica*, 90(1):193–221, 2022. 4, 10

Robert Lucas and Benjamin Moll. Knowledge growth and the allocation of time. *Journal of Political Economy*, 122(1):1–51, 2014. 7

Robert E. Lucas. On the mechanics of economic development. *Journal of Monetary Economics*, 22:3–42, 1988. 3

Erzo G.J. Luttmer. Dynamic urn-ball discovery. Technical report, 2021. 4

Petra Moser. Patents and innovation: Evidence from economic history. *Journal of Economic Perspectives*, 27(1):23–44, 2013. 1, 22

Petra Moser. Patents and innovation in economic history. *Annual Review of Economics*, 8: 241–258, 2016. 1

William D. Nordhaus. *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*. MIT Press, Cambridge, MA, United States, 1969. 1

Richard A Posner. Why there are too many patents in america. *The Atlantic*, 12, 2012. 32, 33

Paul M. Romer. Increasing returns and long-run growth. *Journal of Political Economy*, 94: 1002–1037, 1986. 3, 4

Bhaven Sampat and Heidi L Williams. How do patents affect follow-on innovation? evidence from the human genome. *American Economic Review*, 109(1):203–36, 2019. 2, 5

Frederic M Scherer. Nordhaus' theory of optimal patent life: A geometric reinterpretation. *The American Economic Review*, 62(3):422–427, 1972. 32

Suzanne Scotchmer. Standing on the shoulders of giants: cumulative research and the patent law. *Journal of economic perspectives*, 5(1):29–41, 1991. 5

Terence P. Stewart. *The GATT Uruguay Round: a negotiating history (1986-1992)*, volume 2. Springer, 1993. 33

Has van Vlokhoven. Decomposing the rise in markups. Working paper, 2022. 20

Gerard T. Vondeling, Qi Cao, Maarten J. Postma, and Mark H. Rozenbaum. The impact of patent expiry on drug prices: A systematic literature review. *Applied Health Economics and Health Policy*, 16:653–660, 2018. 23

Heidi L. Williams. How do patents affect research investments? *Annual Reviews of Economics*, 9:441–469, 2017. 1, 2

Appendix

Proof of Proposition 1

Recall Equation (1):

$$dZ_{it} = \begin{cases} (\mu+\theta)dt + \sigma dW_{it} & \text{if } Z_{it} < Z_{jt} \text{ and} \\ \mu dt + \sigma dW_{it} & \text{if } Z_{it} \geq Z_{jt}. \end{cases}$$

By Proposition 5.3.6 of Karatzas and Shreve (1988, p. 303), Equation (1) has a weak solution

$$Z_{it} = Z_{i0} + \int_0^t b_i(Z_{1s}, Z_{2s})ds + \hat{W}_{it},$$

where $b_i(Z_{1s}, Z_{2s})$ is the drift in (1) corresponding to Z_i and $\hat{W}_{it}$ is a 2-dimensional Brownian motion. Since the set $\{s \geq 0 : Z_{1s} = Z_{2s}\}$ almost surely has Lebesgue measure zero (Karatzas and Shreve, 1988, Theorem 2.9.6), the processes (Z_1, Z_2) also weakly solve

$$dZ_{it} = \begin{cases} (\mu + \theta)dt + \sigma dW_{it} & \text{if } Z_{it} < Z_{jt}, \\ \mu dt + \sigma dW_{it} & \text{if } Z_{it} > Z_{jt}, \text{ and} \\ (\mu + \frac{1}{2}\theta)dt + \sigma dW_{it} & \text{if } Z_{it} = Z_{jt}. \end{cases} \tag{1'}$$

From (1), average productivity X has the law of motion

$$dX_t = \begin{cases} (\mu + \frac{1}{2}\theta)dt + \sigma dW_{xt} & \text{if } Z_{1t} \neq Z_{2t}, \text{ and} \\ \mu dt + \sigma dW_{xt} & \text{if } Z_{1t} = Z_{2t}. \end{cases}$$

By the previous argument, X also satisfies the law of motion $dX_t = (\mu + \frac{1}{2}\theta)dt + \sigma dW_{xt}$, as required, by computing dX_t from (1′).

Proof of Proposition 2

By Equation (1), the productivity gap $Y = Z_1 - Z_2$ has the law of motion

$$dY_t = \begin{cases} \frac{1}{2}\theta dt + \sigma dW_{yt} & \text{if } Z_{1t} < Z_{2t}, \\ -\frac{1}{2}\theta dt + \sigma dW_{yt} & \text{if } Z_{1t} > Z_{2t}, \text{ and} \\ 0dt + \sigma dW_{it} & \text{if } Z_{1t} = Z_{2t}, \end{cases}$$

which agrees with the claim, since the drift is precisely $-\frac{1}{2}\theta \text{sgn}(Y_t)$.

It remains to show that X and Y are distributed independently, as the rest of the Lemma follows by direct calculation. Indeed, clearly W^+ and W^- are uncorrelated:

$$\begin{aligned} \mathbb{E}[(W_t^+ - W_s^+)(W_t^- - W_s^-)] &= \tfrac{1}{4}\mathbb{E}[(W_{jt} + W_{kt} - W_{js} - W_{ks})(W_{jt} - W_{kt} - W_{js} + W_{ks})] \\ &= \mathbb{E}[(W_{jt} - W_{js} + (W_{kt} - W_{ks}))(W_{jt} - W_{js} - (W_{kt} - W_{ks}))] \\ &= \mathbb{E}[(W_{jt} - W_{js})^2] - \mathbb{E}[(W_{kt} - W_{ks})^2] = 0. \end{aligned}$$

Independence now follows by joint normality of W_j and W_k.

Proof of Lemma 1

See Fernholz et al. (2013) and Remark 6.5.2 in Karatzas and Shreve (1988, p. 441).

Proof of Proposition 3

Follows immediately by taking limits as $t \to \infty$ of g_t in Lemma 1.

Proof of Proposition 4

See, e.g., Harrison (1985, p. 14).

Proof of Lemma 2

For any fixed time t, consider the cost minimization problem of a final good producer seeking to minimize production cost given a desired output level $\bar{y}_t$:

$$C(\bar{y}_t) = \min_{(y_{ijt})\geq 0} \int_0^{B_t} (p_{i1t}y_{i1t} + p_{i2t}\varepsilon_{it}y_{i2t})di \quad \text{s.t.} \quad y_t = \left[\int_0^{B_t} \left(y_{i1t}^{\beta} + \varepsilon_{it}y_{i2t}^{\beta}\right)^{\alpha/\beta} di\right]^{1/\alpha} \geq \bar{y}_t.$$

where p_{ijt} is the price posted by the producer of intermediate input ij and $B_t > 0$ is the measure of intermediate input varieties in the production of final output.

Profit maximization implies that $C'(y) = 1$, since the final-good price is normalized to one. Cost minimization gives the first-order conditions

$$p_{i1} = \mu_t \frac{1}{\alpha} y_t^{1-\alpha} \alpha y_{i1t}^{\alpha-1}$$

whenever $\varepsilon_{it} = 0$, where μ_t is the multiplier on the constraint $y_t \geq \bar{y}_t$. Since $C'(y_t) = \mu_t$ by the Envelope Theorem, profit maximization implies that $\mu_t = 1$. Rearranging the first-order condition therefore gives $y_{i1t} = y_t p_{i1t}^{-1/(1-\alpha)}$.

If $\varepsilon_{it} = 1$, first-order conditions yield

$$p_{ijt} = y_t^{1-\alpha} \left[y_{i1t}^{\beta} + y_{i2t}^{\beta}\right]^{(\alpha-\beta)/\beta} y_{ijt}^{\beta-1} \quad \text{and} \quad \frac{p_{i1t}}{p_{i2t}} = \left(\frac{y_{i1t}}{y_{i2t}}\right)^{\beta-1}.$$

Therefore,

$$y_{i2t} = y_{i1t} \left(\frac{p_{i1t}}{p_{i2t}}\right)^{1/(1-\beta)}.$$

Substituting this into the first-order conditions gives

$$\begin{aligned}
p_{i1t} &= y_t^{1-\alpha}\left[y_{i1t}^{\beta}+y_{i1t}^{\beta}\left(\frac{p_{i1t}}{p_{i2t}}\right)^{\beta/(1-\beta)}\right]^{(\alpha-\beta)/\beta} y_{i1t}^{\beta-1} \\
&= y_t^{1-\alpha}\left[1+\left(\frac{p_{i1t}}{p_{i2t}}\right)^{\beta/(1-\beta)}\right]^{(\alpha-\beta)/\beta} y_{i1t}^{\alpha-\beta} y_{i1t}^{\beta-1} \\
&= y_t^{1-\alpha}\left[\left(\frac{p_{i1t}}{p_{i1t}}\right)^{\beta/(1-\beta)}+\left(\frac{p_{i1t}}{p_{i2t}}\right)^{\beta/(1-\beta)}\right]^{(\alpha-\beta)/\beta} y_{i1t}^{\alpha-1} \\
&= y_t^{1-\alpha} p_{i1t}^{(\alpha-\beta)/(1-\beta)}\left[p_{i1t}^{-\beta/(1-\beta)}+p_{i2t}^{-\beta/(1-\beta)}\right]^{(\alpha-\beta)/\beta} y_{i1t}^{\alpha-1}.
\end{aligned}$$

Rearranging,

$$p_{i1t}^{(1-\alpha)/(1-\beta)} = y_t^{1-\alpha}\left[p_{i1t}^{-\beta/(1-\beta)}+p_{i2t}^{-\beta/(1-\beta)}\right]^{(\alpha-\beta)/\beta} y_{i1t}^{\alpha-1},$$

which finally yields

$$y_{i1t} = y_t p_{i1t}^{-1/(1-\beta)}\left[p_{i1t}^{-\beta/(1-\beta)}+p_{i2t}^{-\beta/(1-\beta)}\right]^{-(\beta-\alpha)/[\beta(1-\alpha)]}.$$

By symmetry, of course,

$$y_{i2t} = y_t p_{i2t}^{-1/(1-\beta)}\left[p_{i1t}^{-\beta/(1-\beta)}+p_{i2t}^{-\beta/(1-\beta)}\right]^{-(\beta-\alpha)/[\beta(1-\alpha)]},$$

too, as was claimed.

Proof of Proposition 5

Recall that firm $i1$'s profit-maximization problem is given by $\max_p\{y_{i1t}(p,q)(p-c_{i1t})\}$, where $c_{i1t}=w_t/A_{i1t}$ and

$$y_{i1t}(p,q) = \frac{y_t p^{-\frac{1}{1-\beta}}}{\left[p^{-\frac{\beta}{1-\beta}}+q^{-\frac{\beta}{1-\beta}}\right]^{\frac{\beta-\alpha}{\beta(1-\alpha)}}}.$$

Writing $y'_{i1t}=dy_{i1t}/dp$, first-order conditions give the standard inverse-elasticity rule

$$\mu_{i1t} = \frac{p-c_{i1t}}{p} = -\frac{y_{i1t}}{y'_{i1t}p}.$$

Let $R=p^{-\beta/(1-\beta)}+q^{-\beta/(1-\beta)}$. Taking the derivative of the demand curve gives

$$y'_{i1t} = y_t\frac{-1}{1-\beta}p^{-\frac{1}{1-\beta}-1}R^{\frac{\alpha-\beta}{\beta(1-\alpha)}} + yp^{-\frac{1}{1-\beta}}\frac{\alpha-\beta}{\beta(1-\alpha)}R^{\frac{\alpha-\beta}{\beta(1-\alpha)}-1}\frac{-\beta}{1-\beta}p^{-\frac{\beta}{1-\beta}-1}.$$

Therefore,

$$
\begin{aligned}
-\frac{y_{i1t}'}{y_{i1t}} &= \frac{\frac{1}{1-\beta}y_t p^{-\frac{1}{1-\beta}-1}R^{\frac{\alpha-\beta}{\beta(1-\alpha)}} + \frac{\beta}{1-\beta}y_t p^{-\frac{1}{1-\beta}}\frac{\alpha-\beta}{\beta(1-\alpha)}R^{\frac{\alpha-\beta}{\beta(1-\alpha)}-1}p^{-\frac{\beta}{1-\beta}-1}}{y_t p^{-\frac{1}{1-\beta}}R^{\frac{\alpha-\beta}{\beta(1-\alpha)}}} \\
&= \frac{\frac{1}{1-\beta}p^{-\frac{1}{1-\beta}-1} + \frac{\beta}{1-\beta}p^{-\frac{1}{1-\beta}}\frac{\alpha-\beta}{\beta(1-\alpha)}R^{-1}p^{-\frac{\beta}{1-\beta}-1}}{p^{-\frac{1}{1-\beta}}} \\
&= \frac{1}{1-\beta}p^{-1} + \frac{\beta}{1-\beta}\frac{\alpha-\beta}{\beta(1-\alpha)}R^{-1}p^{-\frac{1}{1-\beta}}.
\end{aligned}
$$

This implies that the *reciprocal markup* equals

$$
\frac{1}{\mu_{i1t}} = \frac{p}{p-c_{i1t}} = -\frac{y_{i1t}'p}{y_{i1t}} = \frac{1}{1-\beta} + \frac{\alpha-\beta}{(1-\beta)(1-\alpha)}R^{-1}p^{-\frac{\beta}{1-\beta}}.
$$

Similarly, firm $i2$'s first-order conditions yield

$$
\frac{1}{\mu_{i2t}} = \frac{p}{p-c_{i2t}} = -\frac{y_{i2t}'q}{y_{i2t}} = \frac{1}{1-\beta} + \frac{\alpha-\beta}{(1-\beta)(1-\alpha)}R^{-1}q^{-\frac{\beta}{1-\beta}},
$$

hence

$$
\begin{aligned}
\frac{1}{\mu_{i1t}} + \frac{1}{\mu_{i2t}} = \frac{2}{1-\beta} + \frac{\alpha-\beta}{(1-\beta)(1-\alpha)} &= \frac{2(1-\alpha)+\alpha-\beta}{(1-\beta)(1-\alpha)} \\
&= \frac{2-\alpha-\beta}{(1-\beta)(1-\alpha)} \\
&= \frac{1}{1-\alpha} + \frac{1}{1-\beta}.
\end{aligned}
$$

Going back to the first-order conditions,

$$
\begin{aligned}
\frac{1}{\mu_{i1t}} &= \frac{1}{1-\beta} + \frac{\alpha-\beta}{(1-\beta)(1-\alpha)}\frac{p^{-\frac{\beta}{1-\beta}}}{R} \\
&= \frac{1}{1-\beta}\frac{p^{-\frac{\beta}{1-\beta}} + q^{-\frac{\beta}{1-\beta}}}{R} + \frac{\alpha-\beta}{(1-\beta)(1-\alpha)}\frac{p^{-\frac{\beta}{1-\beta}}}{R} \\
&= \frac{1}{1-\beta}\frac{q^{-\frac{\beta}{1-\beta}}}{R} + \frac{(x\alpha-\beta+1-\alpha)}{(1-\beta)(1-\alpha)}\frac{p^{-\frac{\beta}{1-\beta}}}{R} = \frac{1}{1-\alpha}\frac{p^{-\frac{\beta}{1-\beta}}}{R} + \frac{1}{1-\beta}\frac{q^{-\frac{\beta}{1-\beta}}}{R},
\end{aligned}
$$

and, similarly,

$$
\frac{1}{\mu_{i2t}} = \frac{1}{1-\alpha}\frac{q^{-\frac{\beta}{1-\beta}}}{R} + \frac{1}{1-\beta}\frac{p^{-\frac{\beta}{1-\beta}}}{R}.
$$

Therefore, letting $\rho_1 = \frac{p^{-\frac{\beta}{1-\beta}}}{R}$ and $\rho_2 = \frac{q^{-\frac{\beta}{1-\beta}}}{R}$, for any $j,k \in \{1,2\}$ with $j \neq k$,

$$
\frac{1}{\mu_{ijt}} = \frac{\rho_j}{1-\alpha} + \frac{\rho_k}{1-\beta}. \tag{A1}
$$

Since $\rho_k = 1 - \rho_j$, it follows that

$$\mu_{ijt} = \left[\rho_j(1-\alpha)^{-1} + (1-\rho_j)(1-\beta)^{-1}\right]^{-1}.$$

It remains to establish the existence of such pricing equilibrium. To be more explicit, consider the game between two players, with actions $p_i \in [0, \infty]$ and payoffs given by profits as previously defined. We allow a player's price to equal ∞; this translates to producing a quantity of zero. In fact, given the first-order conditions (A1) above, a player j's best response to its opponent's price must place $\rho_j \in [0, 1]$, since

$$\rho_j = \frac{p_j^{-\frac{\beta}{1-\beta}}}{p_j^{-\frac{\beta}{1-\beta}} + p_k^{-\frac{\beta}{1-\beta}}},$$

where $k \neq j$, is monotonically decreasing in ρ_j, ranging from $\rho_j = 1$ when $p_j = 0$ to $\rho_j = 0$ when $p_j = \infty$.[24] Hence, by (A1), any firm j's best response to given prices for firms 1 and 2 exhibits markups that satisfy $(1-\alpha)^{-1} \leq \mu_j^{-1} \leq (1-\beta)^{-1}$, where μ_j is short for μ_{ijt}. In turn, this implies that best-response prices satisfy $c_j/\beta \leq p_j \leq c_j/\alpha$. Since prices must lie in the compact rectangle $[c_1/\beta, c_1/\alpha] \times [c_2/\beta, c_2/\alpha]$, and the best-response function is clearly continuous, by Brower's Fixed Point Theorem there exists a vector of prices that constitutes a Nash equilibrium of the pricing game above.

Proof of Proposition 6

In the stationary equilibrium of a balanced growth path, there is a constant allocation of labor to research relative to all labor supply, $e^* = \frac{l_{rt}}{l_{rt}+l_{pt}} \in (0, 1)$, which we call "research effort level." In addition, in the balanced growth path, as the stock of blueprints and the population grow at constant rate, there is a stationary distribution of ages for industries.

Consider a candidate stationary equilibrium that features a constant effort level. Suppose that some effort level $e^* \in (0, 1)$ is optimal at some time t, then consider a time $t' \neq t$, the distribution of varieties by age is the same, and the ratio of the stock of blueprints to the labor

[24] This assumes that $p_k \notin \{0, \infty\}$. If $p_k = 0$ then $\rho_j = 0$ whenever $p_j \in (0, \infty]$. If $p_k = \infty$ then $\rho_j = 1$ whenever $p_j \in [0, \infty)$. Finally, if $p_j = p_k = 0$ or $p_j = p_k = \infty$, normalize $\rho_j = \rho_k = \frac{1}{2}$. It will be easy to see that this normalization is without any loss of generality.

supply is fixed. Thus, the distribution of industry revenues by age is the same. Therefore, the fraction of revenues in industries under monopoly (induced by the patent policy) is the same at both points in time. While in other industries, the distribution of productivity differences between firms is described by the probability density $g_{\infty}(.)$. Thus, the average markup level is the same in both t and t'. This implies that the labor share is the same t and t', and, therefore, the effort level e^* is the optimal choice of effort level for time t'.

Thus, to prove that our balanced growth path equilibrium exists, it suffices to show that if there is an effort level $e^* \in (0,1)$ that is the optimal choice for households at some time t, it will be an optimal choice for every t' and therefore consistent with a balanced growth path.

To show this, suppose that the effort level converges to one. As a result, the stock of blueprints increases while the labor supply converges to zero, as the distribution of varieties by age is constant, which implies that the labor share of output is constant. Therefore wages diverge to infinity. If the effort level converges to zero, the stock of blueprints converges to zero, so wages fall to zero. Thus, research effort in a balanced growth path is lower than 1 and higher than 0. In an interior equilibrium, effort $e^* \in (0,1)$ satisfies an optimality condition where the marginal cost of effort, given by the wage rate, equals the present value of profits accrued by a new blueprint multiplied by the productivity of blueprint production (given by the constant c). As wages and the present value of profits vary continuously with the stock of blueprints determined by the effort level, by the intermediate value theorem, there exists an effort level e^* that equates wages to the present value of profits accrued by a new blueprint multiplied by the marginal product of blueprint production. Thus a BGPE exists.

Proof of Proposition 7

We divide the proof into two steps for the cases where $S = 0$ and $S > 0$.

Step 1: If $S = 0$, varying T does not change the growth rate in the balanced growth path.

Proof: If $S = 0$ then $Z_{i1,t_0} = 0$, then the expected productivity in an industry after a length of time $t > 0$ from its creation is $\mu + \max\{0, (t - T)\}\theta/2$. Given that the distribution of productivity by industry ages is constant in the balanced growth path, a change in the T will only shift the average productivity and thus shift the output level in the balanced growth path without changing its growth rate.

Step 2: If $S > 0$, the growth rate in the balanced growth path is strictly decreasing in T.

Proof: In the balanced growth path, the distribution of industry ages is constant, given by F. Let $EZ(t, T)$ be the average Z of the economy at time t under the policy T:

$$EZ(t, T) = \int EZ(a, t, T) dF(a), \tag{3}$$

where $EZ(a, t, T)$ denotes the expected average productivity of industry of a blueprint of age a at time t, which satisfies

$$EZ(t, a, T) = \mu + \max\{0, (t - T)\}\theta/2 + S \times EZ(t - a, T). \tag{4}$$

We want to show that the growth rate of $EZ(T)$ is a decreasing function of T, as it implies that the growth rate in output is also decreasing.

To show that consider two economies in the balanced growth path, denoted by T and T', with $T > T'$, and assume that at some period normalized to 0, $EZ(0, T) = EZ(0, T')$. We need to show that for $t > 0$, then $EZ(t, T) < EZ(t, T')$.

Suppose that $EZ(t, T) > EZ(t, T')$. For the mass of industries that existed before time 0 at period t, they have the same measure $1 - F(t) > 0$ in both economies, and it follows from the assumption that $EZ(0, T) = EZ(0, T')$ that their expected productivity is the same at time 0. Given that a positive fraction $F(T') - F(T)$ of these industries had their productivity growing at a faster rate up to time t because some patents expired in economy T' that did not expire in economy T, then average productivity of this subset of industries is strictly higher in economy T'. The continuity of EZi and the distribution of industry ages F implies that as $t > 0$ becomes small, $1 - F(t)$ converges to one and $EZ(t, T) \leq EZ(t, T')$, a contradiction.

Therefore, $EZ(t, T) \leq EZ(t, T')$ for $t > 0$ and $T > T'$, thus $EZ(t, T)$ is non-increasing in T. That implies that for any $t > 0$, that $EZ(t', T) \leq EZ(t', T')$ for $t' \in (0, t]$. Thus, for startup industries with age $a \in (0, t)$, productivity is equal or higher in economy T'. For the mass of industries that existed before time 0 at period t, they have the same measure $1 - F(a) > 0$ in both economies, and their expected productivity is the same at time 0 (follows from assumption $EZ(0,T) = EZ(0,T')$), given that a positive fraction $F(T') - F(T)$ of these industries had their productivity growing at a faster rate up to time t because some patents expired in economy T' that did not expire in economy T, thus the productivity of this subset of industries is also strictly higher at t. Thus, we have that $EZ(t,T) < EZ(t,T')$.

Proof of Proposition 8

Given wage rate w_t, the utility function $U(c,l)$ of the household implies that the first order condition for the choice of labor supply is as follows:

$$l_t = w_t/(KB_t^{\frac{1-\alpha}{\alpha}}).$$

The wage rate is equal to the labor share times the output per unit of labor. Therefore this functional form, which implicitly assigns a value for leisure that is an increasing function of the stock of blueprints, implies that the labor supply is a linear function of the labor share, which is constant in the balanced growth rate.

Consider a marginal change in the stock of blueprints by $\Delta\%$ and assume that the allocation of labor across blueprints is constant over time (which holds in the balanced growth path), then the production technology implies that aggregate output changes by $\frac{1-\alpha}{\alpha}\Delta\%$. Therefore, the first-order condition for labor supply implies that given constant labor share, the increase in wage rate due to an increase in the stock of blueprints is exactly compensated by a decrease in the denominator.

Proof of Proposition 9

The monopolist will choose to internally replicate symbiotic competition if and only if

$$\int_0^{\infty} \exp(-r\Delta) \left[G_t(\Delta) - E_t[C_{j,t+\Delta}]\right] d\Delta > 0.$$

Note that as $C_{j,t+\Delta}$ is the expected profit flow the monopolist expects that the manager obtains if he runs his own firm as one of the firms under a duopoly in time $t + \Delta$, this expected profit is always strictly positive, and it grows according to the growth in productivity of the sector in duopoly (which grows at the expected rate $\mu + \theta/2$, the same as the expected growth rate of productivity under monopoly with two divisions) and according to the growth of the economy (as described by the parameter y_t).

Therefore, $C_{j,t+\Delta} > 0$ for every $\Delta \geq 0$ while $G_t(\Delta)$ converges to zero as Δ converges to zero. Thus, there exists a time interval $[0, \overline{\Delta}]$, with $\overline{\Delta} \in (0, \infty)$, such that the monopolist expects negative profits by hiring the manager. Note that $\overline{\Delta} > 0$ cannot be infinity: as Δ diverges to infinity, the faster rate of productivity increase implies that the increase in monopoly profits from shifting to the two division-system a relative increase in profits that is infinitely many times higher than under single division-system. Since profits under a monopoly are higher than profits under a duopoly given the same industry-average technology, therefore the increase in expected monopoly profits of a shift from one division to two divisions becomes higher than profits under competition for $\Delta > 0$ large enough. Therefore, profits are lower than the costs of the manager from $[0, \overline{\Delta}]$ and higher than the costs of the manager during the period $[\overline{\Delta}, \infty)$. Thus, if the discount rate $r > 0$ is high enough, the monopolist never finds profit-maximizing to replicate the effects of symbiotic competition under duopoly.