# Technology Choice, Spillovers, and the Concentration of R&D

Todd Lensman*

November 25, 2024

[Please click here for an updated draft]

**Abstract**

The direction of innovation shapes both current technologies and future innovation opportunities, as firms acquire expertise and create public knowledge through discovery. But how do firms choose which technologies to develop? Do they ever fail to exploit new technological paradigms? I build a new model of innovation and firm dynamics to study a novel link between market structure, the direction of innovation, and economic growth: Expertise in a current technology gives incumbents a comparative advantage at innovating it relative to entrants, who instead favor a new technology with higher growth potential. Each firm's innovation decisions influence others through knowledge spillovers, so the initial market structure can affect the long-run direction of innovation. Concentrating R&D resources in a small number of firms allows faster accumulation of expertise. This raises growth when all firms innovate the same technology. But it can lower growth when firms face a technology choice, amplifying the influence of incumbents and potentially delaying or preventing the emergence of the new technology. I provide empirical evidence for the theory using data on firm patenting and R&D expenditures. I also show that it explains the historical development of mRNA vaccines, and I explore its implications for the highly concentrated innovation of artificial intelligence.

**JEL Classification:** L16, L25, O31, O33, O41

**Keywords:** economic growth, innovation, directed technological change, firm dynamics, technology paradigms

*Massachusetts Institute of Technology, Department of Economics (email: tlensman@mit.edu).
I thank my advisors, Daron Acemoglu, Michael Whinston, and Jacob Moscona, for their guidance and encouragement throughout this project. I also thank María Ballesteros, Ricardo Caballero, Jean Tirole, Iván Werning, Christian Wolf, and participants at MIT seminars and lunches for their helpful suggestions. I gratefully acknowledge financial support from the National Science Foundation's Graduate Research Fellowship Program.

# 1 Introduction

Rapid progress in artificial intelligence (AI) has sparked intense debate about how firms are developing this technology and whether governments should intervene. Two issues are prominent: First, AI is a general-purpose technology that can be applied to a wide variety of tasks, from predicting the structure of proteins to personalizing online ads. Firms developing AI must choose which applications to target, but there is broad concern that the resulting *direction of innovation* may be not be in society's best interest.[1] Second, many of these decisions are made by just a handful of large, incumbent technology firms. These firms may face different innovation incentives than society or smaller entrants, and they also control vast research and development (R&D) resources that lend them influence over the direction of innovation. Figure 1 shows that the *concentration of R&D* is a recent and growing macro phenomenon. In the past decade, the share of US R&D expenditures accounted for by the top five US public firms has nearly doubled to 22.5%. This trend coincides exactly with the rise of the "Big Five" technology firms to the top of the R&D rankings.

Concerns about the direction of innovation and the concentration of R&D are salient for AI, but they apply to many industries with large incumbents, including the aerospace and defense, automotive, and pharmaceutical industries. They also raise several fundamental questions about the relationship between market structure and innovation:

(i) What incentives drive the innovation direction of incumbents, and how do they differ from the incentives faced by entrants or a welfare-minded social planner?

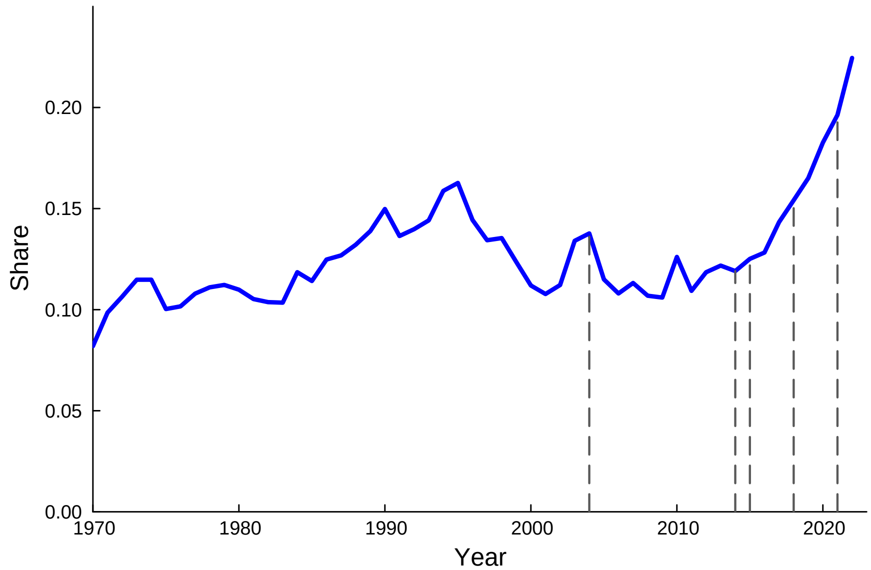
(ii) How do firms' decisions collectively determine the aggregate direction of innovation?

(iii) How does the concentration of R&D within large firms affect aggregate innovation?

The central argument of this paper is that our current answers to these questions are incomplete. They overlook how a basic feature of the innovation process — the accumulation of knowledge — drives incumbents and entrants to develop different technologies, while linking these decisions through spillovers to determine the aggregate direction of innovation. I build a new model of directed innovation and firm dynamics to clarify this mechanism. The model reveals how market structure influences innovation not just through competition, but because firms with different accumulated expertise choose to develop different technologies. Their discoveries create new innovation opportunities for all firms, so that an industry's market structure can affect its long-run direction. The direction may be socially inefficient precisely because of these spillovers. And the concentration of R&D plays a critical role by shaping

[1] For example, see Acemoglu (2021), Brynjolfsson (2023), and un.org/sg/en/content/sg/statement/2024-01-17/secretary-generals-special-address-the-world-economic-forum-delivered.

Figure 1: Share of US R&D Expenditures, Top 5 US Public Firms 1970-2022




*Notes:* The vertical lines mark the first year Microsoft (2004), Google (2014), Amazon (2015), Apple (2018), and Meta (2021) each became one of the top five R&D firms. All except for Microsoft have remained in the top five after entry; Microsoft has continuously appeared since 2011. Data on firm R&D expenditures come from Compustat North America, and data on aggregate US R&D expenditures come from the National Science Board.

innovation incentives and the market structure itself.

I provide evidence for the theory through an empirical analysis of firm patenting. Firms generally build on their own discoveries, and incumbents with greater patenting experience are slower to innovate emerging technologies. I also discuss a case study on the development of mRNA vaccines, where my theory offers a simple explanation for firm innovation decisions. This case demonstrates the advantages of decentralized R&D in promoting the exploration of new technologies, contrasting sharply with the current pattern of innovation in AI.[2]

**Model.** I formulate the theory in an endogenous growth model with two key ingredients. First, there are two technologies used by a large number of firms in production. Each firm can employ a team of scientists to innovate and raise its technology-specific qualities (productivities). But innovations for one technology are not useful for the other, so firms must choose *which* technology to innovate. In this sense, the technologies represent not just production processes, but *technological paradigms* that structure how firms can produce and innovate (Dosi, 1982). The concentration of R&D across firms is controlled by the number of scientists a firm

[2] See Klinger et al. (2020), Jurowetzki et al. (2021), and Ahmed et al. (2023) for suggestive evidence that growing concentration of AI research in corporations is driving falling diversity in research topics.

can manage. With the aggregate number of scientists fixed, larger team sizes allow for faster innovation but within fewer firms.

Second, innovations produce knowledge for each technology that raises the productivity of research directed toward it. Some knowledge accumulates publicly and can be used by any firm, a key source of spillovers (Romer, 1990). But crucially I assume that knowledge also accumulates *within the firm*, embodied in its qualities. For example, firms may specialize on particular research lines within technologies, or firms may work as a means to internalize knowledge spillovers between scientists. As a result, a firm's current research productivities depend directly on its past innovation decisions, and firm boundaries matter in the growth process. This assumption is consistent with recent evidence from the auto industry (Aghion et al., 2016). I provide evidence from a broader range of industries in Section 5.

All firms initially innovate for an "old" technology, and incumbent firms exit at random and are replaced by new entrants. As in existing models of the firm-size distribution (e.g., Luttmer, 2007), the combination of innovation and stochastic entry and exit generates heterogeneity — incumbents with a long record of innovation have higher qualities than recent entrants. I then consider the unanticipated arrival of a "new" technology. The new technology allows faster growth than the old technology in the long run, but with little accumulated knowledge it offers slower growth in the short run. In the context of AI, the old technology might represent integration with existing products like office productivity software or internet search, while the new technology might represent more radical applications like automated drug discovery or adaptive learning. Firms must choose which technology to innovate; the bulk of the paper analyzes their equilibrium choices.

**Results.** The model yields four main insights. First, greater concentration of R&D raises the economy's growth rate along its balanced growth paths, in which all firms innovate the same technology (Proposition 1). This result establishes a baseline role for *the firm* in economic growth even without a technology choice: Larger firms can focus more R&D resources on their research lines, developing greater expertise that helps them innovate even more quickly.[3] It also suggests an optimistic interpretation of Figure 1, that rising concentration of R&D might simply allow faster innovation by the largest firms.

To assess whether this holds when firms face a technology choice, I consider the introduction of the new technology. The second main result shows that innovation decisions are both *path-dependent* and *forward-looking* at the firm level (Proposition 2): Firms have incentives to continue innovating a technology in which they have developed expertise, but they also

[3] This result also offers a novel foundation for Schumpeter's (1942) claim that larger firms boost growth, but independent of his original justification that size enables greater surplus extraction in the goods market.

consider whether to pursue faster long-run growth by innovating the new technology. The knowledge accumulation mechanism thus generates endogenous comparative advantage *in innovation*, predicting that experienced incumbents will continue innovating the old technology while younger firms and entrants will embrace the new one. For example, the "Big Five" technology firms from Figure 1 have natural incentives to incorporate AI into their existing products, even if other applications hold greater long-run promise.

This explanation for heterogeneous innovation incentives differs fundamentally from most work on market structure and innovation. A large literature following Arrow (1962) studies how competition for monopoly rents can incentivize incumbents to innovate more or less than entrants, depending on whether innovation cannibalizes incumbents' existing products or allows them to escape competition from entrants.[4] A distinct literature following Henderson and Clark (1990) emphasizes that incumbent firms may have an *absolute* disadvantage at innovating new technologies because of organizational rigidities.[5] As a result, the arrival of new technologies that compete with old ones can trigger the failure of established incumbents. The knowledge accumulation mechanism I identify is independent of these effects. I show this by deliberately constructing the model to rule out competition between firms or technologies in the goods market. I also assume that incumbents do not have an absolute disadvantage at innovating the new technology when it arrives. Rather, the key feature in my model is their *comparative* disadvantage at innovating the new technology.

I next turn to the aggregation of firm innovation decisions in equilibrium. The third main result shows that the economy's initial *market structure* (firm-quality distribution) can affect the aggregate technology choice in the long run. As in past work on technology paradigms, spillovers through public knowledge encourage firms to innovate the same technology, so that initial innovation decisions can permanently "tip" the economy toward one technology or the other (Arthur, 1989; Acemoglu et al., 2012). But in my model firms' innovation decisions depend on their expertise, so the direction of this effect depends critically on the initial market structure. For example, an industry dominated by large incumbents may fail to exploit a new technology not just because the incumbents favor the old technology, but because they generate spillovers that induce other firms to join them. These complementarities introduce substantial technical challenges, which I resolve by defining, characterizing, and proving the existence of *monotone equilibria*. Monotone equilibria are tractable and allow for rich interactions between firms, reflecting the patterns of innovation found in the case study and the empirical analysis.

[4]The "replacement effect" is formalized by Arrow (1962) and Reinganum (1983), while the "escape-competition effect" is formalized by Gilbert and Newbery (1982), Aghion et al. (2001), and Aghion et al. (2005). Acemoglu and Cao (2015) explore the implications of the replacement effect for incremental and radical innovations, while Igami (2017) provides an empirical analysis of these issues in the hard drive disk industry.

[5]See also Henderson (1993) and Christensen (1997).

The final main result reconsiders the concentration of R&D when firms face a technology choice (Proposition 5). Greater concentration exacerbates both the path-dependent and forward-looking forces: It allows incumbents to accumulate more expertise for the old technology before the new one arrives, but it also enhances the new technology's growth advantage. The former dominates when the discount rate is sufficiently high. An increase in R&D concentration can then induce "lock-in" to the old technology, and a simple quantitative example in Section 6 shows that this holds for empirically reasonable parameter values. Lock-in may be inefficient because of knowledge spillovers, and in general a social planner would always transition to the new technology more quickly than in equilibrium (Proposition 6). These findings suggest an alternative interpretation to Figure 1: The rising concentration of R&D may drive lower diversity in research as firms follow in the footsteps of large incumbents, potentially *reducing* long-run growth as alternative directions remain unexplored.

**Evidence.** To provide evidence for the theory, I conduct a case study and an empirical analysis of US patent data. The case study traces the development of mRNA vaccine technology, which was critical to the recovery from the COVID-19 pandemic of 2020-2023. Historical accounts emphasize that small biotechnology firms were largely responsible for innovating this "new" technology, while large pharmaceutical incumbents with expertise in conventional ("old") vaccines decided against it (Dolgin, 2021); these decisions are also captured in the patent record. Competition- and organization-based theories of innovation incentives have difficulty explaining this pattern, because conventional and mRNA vaccines are often not competing for applications. My theory instead offers a simple explanation based on comparative advantage in innovation, and it rationalizes why several incumbent firms have recently begun to explore mRNA technology. This case also suggests the potential stakes of R&D concentration: Had incumbents exercised greater control over R&D resources, we might have been left without a crucial tool to fight the COVID-19 pandemic.

To show that the lessons of the case study generalize, I study firm innovation decisions using a panel of US patents matched to US public firms over 1980-2021 (Arora et al., 2024). I establish three facts consistent with the theory. First, a firm's current patenting is highly correlated with its previous patenting, controlling for R&D expenditures and a variety of other determinants of innovation. This provides evidence that technological knowledge accumulates within firms, raising their research productivities. Second, after clustering patents according to the new technologies identified by Kalyani et al. (2023), I find that a firm's patenting for a technology is better predicted by previous patenting *within that technology* than by patenting in general. This suggests that knowledge is not only cumulative within firms, but technology-specific — two critical assumptions of the theory. Finally, I show that incumbents with greater patenting experience innovate substantially *less* for new technologies than less-experienced firms, though this gap shrinks as the technologies mature. This finding supports the theory's main prediction that experienced incumbents should be reluctant to embrace new technologies, and it reflects the same pattern of innovation exhibited in the case study. Thus the empirical analysis broadly supports both the assumptions and implications of the theory, and it also provides parameter estimates used to calibrate the quantitative example in Section 6.

**Related Literature.** I contribute to an expansive literature on endogenous growth (Romer, 1986, 1990; Grossman and Helpman, 1991; Jones, 1995), particularly work relating growth to market structure or firm dynamics (Aghion and Howitt, 1992; Klette and Kortum, 2004; Aghion et al., 2005; Acemoglu and Cao, 2015; Akcigit and Kerr, 2018; Akcigit and Ates, 2021, 2023). This literature primarily addresses growth within a technological paradigm and views market structure through the lens of competition, showing how the associated "replacement" and "escape-competition" effects shape the *rate* of innovation. I instead consider a choice between technological paradigms, and I highlight a distinct role for market structure to affect the *direction* of innovation through knowledge accumulation. I find that conditions favorable for growth within a technological paradigm (e.g., concentrated R&D) can reduce growth by hindering the emergence of a new one. My framework also builds on models of the firm-productivity distribution that imply a role for *the firm* in accumulating knowledge (Luttmer, 2007; Lucas and Moll, 2014; Benhabib et al., 2021; König et al., 2022).

More closely related is a literature on paradigms and increasing returns to scale in the development and adoption of new technologies. Early work emphasizes how firms typically innovate within established paradigms, limited by bounded rationality or organizational constraints from exploring new ones (Nelson and Winter, 1982; Dosi, 1982).[6] Arthur (1989) first discusses how technological lock-in can arise from complementarities in adoption decisions. Farrell and Saloner (1986) and Katz and Shapiro (1986) develop similar ideas in the theory of network effects, and they explore how owners of the underlying technologies can strategically influence adoption. I consider network effects in the context of innovation, focusing on how heterogeneity across firms (market structure) affects the equilibrium choice of a paradigm.

A growing body of work integrates technological paradigms into models of economic growth. Acemoglu (2011) shows that creative destruction within paradigms can dissuade firms from exploring alternatives, leading to too little technological diversity in equilibrium. I show how loss of diversity can arise instead through knowledge accumulation that raises the opportunity cost of exploration, with a critical role played by market structure and the concentration of R&D.

[6] This work was itself inspired by a literature on *scientific* paradigms following Kuhn (1970). See Brock and Durlauf (1999) for a formalization emphasizing the role of conformity effects.

Few papers in this literature consider how different innovation incentives for incumbents and entrants impact the aggregate technology choice. Acemoglu et al. (2016) build a quantitative model of directed innovation and firm dynamics to study the speed of the clean transition under various policy regimes. They assume firms become more productive at innovating technologies based on their past experience, but this model is explicitly designed to aggregate: As in Klette and Kortum (2004), knowledge accumulates only at the product level, so that firm boundaries and market structure play no role in the innovation process. The most closely related paper to mine is a contemporaneous contribution by Aghion et al. (2024). They consider an innovation process similar to that of Acemoglu et al. (2016) and break aggregation by assuming that firms of different ages face different credit constraints. As a result, firm boundaries and market structure affect innovation only insofar as credit constraints bind, and they use their model to quantitatively explore how credit conditions can affect clean innovation.

I also contribute to the broad literature on market structure and innovation. In addition to the seminal contributions discussed above, recent work assesses how incumbents can block innovation by competitors through acquistions (Cunningham et al., 2021), defensive patenting (Argente et al., 2020), and pre-emptive hiring of inventors (Akcigit and Goldschlag, 2023). The knowledge accumulation mechanism I study works independently of competitive pressures, and hinges instead on how complementarities between firms lend incumbents influence over the direction of innovation. In a potentially surprising contrast, I find that incumbents can reduce long-run growth *precisely by innovating according to their expertise*, generating knowledge spillovers on other firms that raise the opportunity cost of exploring new technologies.

Finally, my empirical analysis contributes to a large literature using panel data to estimate the determinants and effects of innovation at the firm level (e.g., Griliches, 1998; Bloom et al., 2013; Kalyani et al., 2023; see Hall et al., 2010 for a review), particularly recent work on directed innovation and path dependence in clean and dirty technologies (Dechezleprêtre et al., 2014; Aghion et al., 2016; Dugoua and Gerarden, 2023). I provide evidence for path dependence from a broader set of industries. I also document the novel result that incumbents are slow to innovate emerging technologies, but do so after other firms make early progress.

**Outline.** The rest of this paper is organized as follows: Section 2 presents the mRNA case study. Section 3 sets up the baseline model of directed innovation and firm dynamics and characterizes its balanced growth paths. Section 4 analyzes equilibria after the introduction of the new technology. Section 5 describes the empirical analysis, and Section 6 presents a simple quantitative example. Section 7 concludes.

# 2 Case Study: mRNA Vaccines

In this section, I provide a brief case study of the development of mRNA vaccines. This case study is instructive in three ways. First, it provides a clear example of the economic setting described by the theory, with distinct old and new technologies that produce different goods and build on different bodies of knowledge. Second, evidence from historical accounts and the patent data shows that incumbent and entrant firms innovated different technologies. I argue that the competition- and organization-based theories of innovation incentives cannot explain this pattern, which is instead consistent with the knowledge accumulation mechanism outlined in my theory. Finally, this case also offers a sharp contrast to the concentration of R&D in AI: Pharmaceutical research is broadly decentralized across many firms, universities, and public and private research institutes. I argue that life-saving mRNA vaccines emerged *precisely because* of this decentralization, underscoring the substantial risks to concentrating R&D among a small number of firms in any industry.

## 2.1 Background

The spread of COVID-19 in early 2020 triggered one of the worst pandemics in a century and the sharpest economic contraction in the post-war period. Within a year, COVID-19 caused over half a million deaths in the United States, vaulting to third among the country's leading causes of mortality.[7] A combination of pandemic-driven uncertainty and strict public health interventions drove the unemployment rate in the United States to a high of nearly 15%, while real output fell by 7.5% in the second quarter of 2020 alone. Given the severity of these initial impacts, the subsequent recovery is all the more remarkable: In just over three years, rapid growth in productivity, employment, and business creation returned the unemployment rate to its pre-pandemic level and real output to its pre-pandemic trend (de Soyres et al., 2024). The infection rate and the risk of hospitalization or death from COVID-19 have also fallen dramatically from their peaks in 2021. A variety of policies and treatment advances supported these outcomes, but one critical innovation stands out: vaccines.

The race to develop vaccines for COVID-19 began at the onset of the pandemic, spearheaded by pharmaceutical companies with public backing through the US government's Operation Warp Speed program. The first two vaccines to receive approval were produced by biotech firm Moderna and a joint venture between pharmaceutical giant Pfizer and biotech firm BioNTech. These vaccines proved exceptionally effective, and they remain the most widely

[7] See covid.cdc.gov/covid-data-tracker for COVID-19 statistics and Murphy et al. (2021) for causes of mortality. See fred.stlouisfed.org for all US macroeconomic statistics.

adopted COVID-19 vaccines in the United States.[8] Their underlying technology is particularly notable: Both vaccines are based on a novel technique for producing immune resistance that differs fundamentally from the mechanism used by conventional vaccines. Conventional vaccines stimulate resistance by confronting the immune system with a weakened version of a pathogen or one of its constituent proteins, both grown in a lab. The Moderna and Pfizer-BioNTech vaccines instead encode instructions for the production of a protein using a genetic material called *messenger RNA* (mRNA). *mRNA vaccines* ferry these instructions into human cells, where they are used to produce the protein en masse to stimulate immune resistance (Hedestam and Sandberg, 2023).

The contrast between conventional and mRNA vaccines offers a vivid illustration of my theory. This case features clearly identifiable old (conventional) and new (mRNA) technologies, both of which reflect the basic features of technological paradigms: Conventional and mRNA vaccine technologies are not just individual products, but sets of techniques used to *produce* and *improve* a variety of products (vaccines). Each technology also required substantial innovation before a viable product could be produced. For example, the practice of conventional vaccination dates back over a millenium, but the production of vaccines for many diseases was not possible before the development of microbiology in the late 19th century (Kinch, 2018). By contrast, mRNA was first discovered only in 1961, and initial experiments suggesting its therapeutic potential were conducted in the early 1990s. The final breakthrough enabling vaccine development came in experiments by Katalin Karikó and Drew Weissman in 2005, for which these researchers were awarded a Nobel Prize in 2023 (Karikó et al., 2005; Hedestam and Sandberg, 2023). As this discussion indicates, the bodies of knowledge for conventional and mRNA vaccine technologies are largely distinct.[9] A firm interested in innovating a new vaccine then faces the important choice of *which* technology to research.[10]

## 2.2 mRNA Innovation

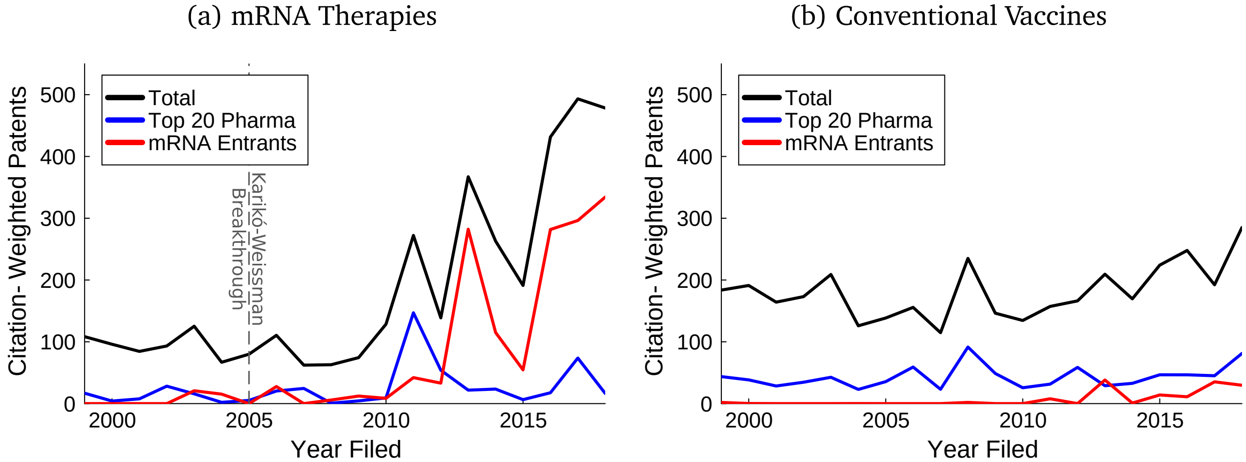
In the case of mRNA vaccines, incumbent and entrant firms made different innovation choices. Figure 2 displays the number of patents related to mRNA therapies and conven-

[8] See ourworldindata.org/grapher/covid-vaccine-doses-by-manufacturer?country= USA

[9] The patent data provide additional evidence that knowledge is technology-specific. Using the sets of mRNA therapy and conventional vaccine patents described below, I compute the the frequency with which patents in one set cite patents in each set. On average, a share 0.14 of citations by an mRNA patent are made to other mRNA patents, while this "within-technology" citation share is 0.39 for conventional vaccine patents. By contrast, on average only a share 0.03 of citations by an mRNA patent are made to conventional vaccine patents, while this "cross-technology" citation share is 0.01 for conventional vaccine patents. The low cross-technology shares suggest that the knowledge produced by one technology is not generally useful for the other.

[10] In reality firms can research both, and the patent data show that they often do. But they must choose between the two technologies at least for the marginal unit of R&D expenditure, and any non-convexities in the required expenditures can make this choice more discrete.

Figure 2: Citation-Weighted Patents, 1999-2018




*Notes:* Figures 2(a) and 2(b) respectively include all mRNA therapy and conventional vaccine patents filed at the US Patent and Trademark Office (USPTO) over 1999-2018 and granted before 2023. I halt both figures in 2018 to mitigate truncation issues from the lag between patent filing and publication. The vertical dashed line in Figure 2(a) marks the year of the Karikó-Weissman breakthrough (Karikó et al., 2005).

tional vaccines that were filed (and eventually granted) in the United States between 1999 and 2018.[11] To more accurately reflect its scientific contribution, each patent is weighted by the number of forward citations from future patents, controlling for the time horizon after publication. The black line in each figure represents the total number of patents filed each year, including patents assigned to individual researchers, firms, universities, public research organizations, and independent research institutes. The blue line represents the subset of patents assigned to the top twenty pharmaceutical companies by revenue in 2023, while the red line represents patents assigned to just four entrant firms whose founders played a key role in the development of mRNA technology: Moderna, BioNTech, the pharmaceutical firm CureVac, and the now-defunct biotech firm RNARx (Hedestam and Sandberg, 2023). All entrants were founded after 2000 specifically to commercialize mRNA therapies.

Figure 2 demonstrates two facts. First, a substantial share of research on mRNA therapies and conventional vaccines is performed outside of the identified incumbent and entrant firms, reflecting the broader decentralization of research in the pharmaceutical industry (Scott Morton and Kyle, 2011). Early contributions to mRNA technology were made primarily by researchers at universities and independent research institutes; Figure 2(a) shows that researchers outside of large pharmaceutical firms or the leading mRNA therapy firms remain

[11] I consider a broader set of patents than just those for mRNA vaccines, because the underlying technology can also be applied to treat a variety of genetic conditions and cancers. See Appendix D for a full description of the patent data, citation weights, and procedure for identifying mRNA therapy and conventional vaccine patents.

influential. Second, the small set of entrant firms played a much bigger role in the innovation of mRNA therapies than the larger set of incumbent pharmaceutical firms. A recent historical account by Dolgin (2021) emphasizes that this was an active decision by many of these incumbents: "In the 1990s and for most of the 2000s, nearly every vaccine company that considered working on mRNA opted to invest its resources elswhere." This remained true even after the final technical barrier to mRNA therapy was resolved in 2005.[12] The entrant firms noted above were responsible not only for much of the foundational basic research on mRNA technology, but also the initial therapy development: CureVac conducted the first clinical trial for an mRNA vaccine in 2017, and Moderna followed quickly to test mRNA vaccines against the Zika virus and two strains of the avian influenza virus.[13] Throughout the time horizon, incumbent pharmaceutical firms remained more active in conventional vaccine innovation than in mRNA therapy innovation (Figure 2(b)).

Why were incumbents so reluctant to invest in mRNA technology, while entrants pushed it from the laboratory to pharmacy shelves? I argue that competition-based theories of innovation incentives cannot explain this pattern, because the mRNA and conventional vaccine technologies are often *not competing* for uses. Each type of vaccine has its own advantages and disadvantages that make it suitable for different applications. For example, mRNA vaccines are often faster to develop and easier to produce than conventional vaccines, but to maintain efficacy they must be transported and stored at below-freezing temperatures. This limits their use in many contexts, where conventional vaccines may be more suitable (Gote et al., 2023). Firms developing mRNA therapies have also focused on conditions without existing vaccines and/or treatments, including infectious diseases like HIV and avian flu, genetic diseases like cystic fibrosis, and a variety of cancers.[14] These considerations limit the extent to which mRNA therapies cannibalize or "steal business" from existing therapies, the standard explanation for why incumbents might be reluctant to innovate relative to entrants (Arrow, 1962; Reinganum, 1983). Moreover, in the wake of the technology's success during the COVID-19 pandemic, several incumbent pharmaceutical firms have begun developing mRNA therapies.[15] This observation sheds doubt on organizational theories that suggests incumbents did not innovate because they *could not* innovate (Henderson, 1993).

My theory instead offers a simple explanation based on knowledge accumulation within firms: Incumbent pharmaceutical firms had already developed *expertise* in other areas of re-

[12] In a recent interview, Weissman recalled that after their 2005 breakthrough he "told [Karikó] our phones are going to ring off the hook. But nothing happened. We didn't get a single call." (Yu, 2021)

[13] See Alberer et al. (2017) and trials NCT03014089, NCT03076385, and NCT03345043 at ClinicalTrials.gov.

[14] See the product pipelines for Moderna (modernatx.com/en-US/research/product-pipeline) and BioNTech (biontech.com/int/en/home/pipeline-and-products/pipeline.html).

[15] See the product pipelines for Merck (merck.com/research/product-pipeline), Pfizer (pfizer.com/science/drug-product-pipeline), and Roche (roche.com/solutions/pipeline).

search (including conventional vaccines) by the time mRNA technology emerged in the 1990s. In line with the quote from Dolgin (2021) above, these firms found it more valuable to continue innovating in those areas than fund early work on mRNA technology. With no such expertise, the entrants instead pursued research for a less developed technology with greater long-run promise. Only after learning from years' worth of innovation by entrants did the incumbent pharmaceutical firms begin seriously investing in mRNA technology. The empirical analysis in Section 5 shows that this pattern of innovation by incumbents and entrants generalizes to many new technologies. A key contribution of the model in Section 3 is to show that it naturally arises when firms accumulate expertise through innovation.

However, the model also shows that this pattern of innovation holds only if entrants have enough R&D resources to push the new technology forward. Otherwise, the entrants may be unable to make sufficient progress before it becomes more valuable for them to abandon the new technology and join the incumbents in innovating the old one. This observation highlights an advantage to decentralized R&D ecosystems like that in the pharmaceutical industry: Although incumbents almost universally decided against funding mRNA therapy research, entrants were able to develop the technology to the point that they could distribute highly effective COVID-19 vaccines within a year of the initial outbreak. This decentralization contrasts sharply with current trends in AI, where the "Big Five" technology firms from Figure 1 deploy the vast majority of R&D resources while hiring researchers from academia and startups at a growing rate (Ahmed et al., 2023).[16] The associated risks may be substantial: Incumbents may leave socially valuable technologies undeveloped, and entrants may simply choose to join them instead of independently developing other technologies. The model developed in the next two sections describes why and when this might occur.

# 3 Model

This section describes the baseline model of directed innovation and firm dynamics. I set up the model and define an equilibrium in Section 3.1, I solve the "static block" of the model in Section 3.2, and I define and characterize balanced growth in Section 3.3.

## 3.1 Setup

**Consumption.** The economy is deterministic and exists in continuous time, populated by a mass $L > 0$ of *workers*, a mass $S > 0$ of *scientists*, and a mass $N \in (0, 1]$ of *entrepreneurs*. To focus attention on firm innovation decisions, I keep the demand side of the economy as

[16] See also https://www.nytimes.com/2024/06/13/opinion/big-tech-ftc-ai.html.

simple as possible: All agents have linear preferences over a unique consumption good, with common discount rate $\rho > 0$. The economy admits a representative consumer who evaluates consumption streams $[C(t)]_t$ with the utility function

$$\int_0^\infty \exp(-\rho t)\, C(t) dt. \tag{1}$$

Workers inelastically supply one unit of labor at wage $w_L(t)$, while entrepreneurs and scientists own all firms and earn profits as described below. All agents can risklessly save and borrow against the total value of firms $\mathcal{A}(t)$ at the equilibrium interest rate $r(t)$. Given the path of interest rates $[r(t)]_t$, the representative consumer solves a standard consumption-savings problem to maximize her utility (1) subject to her budget constraint and no-Ponzi condition:

$$\dot{\mathcal{A}}(t) \leq w_L(t)L + r(t)\mathcal{A}(t) - C(t), \tag{2}$$

$$0 \leq \lim_{t\to\infty} \mathcal{A}(t)\exp\left(-\int_0^t r(s)ds\right). \tag{3}$$

I take the final (consumption) good as the numeraire. The main simplifying assumption here is that preferences are linear, so that in equilibrium the interest rate is fixed at $r(t) = \rho$. With less elastic intertemporal preferences, the interest rate would vary along a technological transition and introduce additional dynamics into firm innovation decisions. This general equilibrium effect is potentially interesting but obscures the more fundamental innovation incentives at the heart of my analysis.

**Production.** A competitive firm produces the final good by combining labor supplied by workers with intermediates. Intermediates come in two types $\theta \in \{A, B\}$ and are supplied by a unit measure of monopolistic *incumbent* firms; I refer to each type $\theta$ as a *technology*. Each incumbent owns one intermediate for each technology $\theta$ with an endogenous and firm-specific *quality* $q_\theta(t)$. An incumbent is fully characterized by its vector of qualities $q \equiv (q_A, q_B)$ at each time, and I let $F(q,t)$ denote the distribution of qualities across incumbents at time $t$. In the context of the mRNA case study, the technologies $A$ and $B$ correspond to conventional and mRNA vaccines, and the qualities $q$ represent a firm's ability to produce effective vaccines with each technology.

Given labor input $L(t)$ and intermediate inputs $x_A(q,t)$ and $x_B(q,t)$ from each firm with qualities $q$, final output is

$$Y(t) \equiv \frac{1}{1-\beta}\left(\int \sum_{\theta\in\{A,B\}} q_\theta^\beta x_\theta(q,t)^{1-\beta} dF(q,t)\right) L(t)^\beta. \tag{4}$$

The final producer chooses the inputs $L(t)$ and $[x_\theta(q,t)]_{\theta,q}$ at each time to maximize its profits, taking as given the wage for workers $w_L(t)$ and the intermediate prices $[p_\theta(q,t)]_{\theta,q}$. The wage $w_L(t)$ is set competitively to clear the market for production labor:

$$L(t) = L. \tag{5}$$

The intermediate prices $[p_\theta(q,t)]_{\theta,q}$ are chosen by firms to maximize profits as described below. Each unit of an intermediate is produced using $\gamma > 0$ units of final output. All remaining output is used for consumption, yielding the market-clearing condition

$$Y(t) = C(t) + \gamma \int \sum_{\theta\in\{A,B\}} x_\theta(q,t) dF(q,t). \tag{6}$$

This production structure is used frequently in models of endogenous growth with quality upgrading, though I introduce the distinction between two sets of intermediates $\theta \in \{A,B\}$.[17] Its key features are additive separability across intermediates in the final production function (4) and the absence of any labor reallocation between intermediates. As a result, the final producer's demand curve for each intermediate is independent of the qualities and prices of all others, $x_\theta(q,t) = q_\theta p_\theta(q,t)^{-\frac{1}{\beta}} L$. The flow profits for firms inherit these properties, as each firm sets the prices of its intermediates $p_A(q,t)$ and $p_B(q,t)$ to maximize profits $\pi$ given its qualities $q$:[18]

$$\pi(q) \equiv (q_A + q_B)\,\bar{\pi}, \quad \text{where} \quad \bar{\pi} \equiv \max_p (p-\gamma)\, p^{-\frac{1}{\beta}} L. \tag{7}$$

This expression demonstrates that the economy features no competition between firms or substitution between technologies that could affect innovation incentives. Competition is ruled out because each firm's profits do not depend on other firms' qualities, so firms will not innovate to steal business or preempt competition from rival firms. Similarly, the profits earned by technology $A$ intermediates do not depend on the qualities of technology $B$ intermediates. Such dependence would be natural if technologies $A$ and $B$ instead produced distinct goods

[17] For example, see Howitt (1999) and Acemoglu et al. (2006).

[18] Flow profits are also linear and symmetric in the qualities $q$. Linearity is essentially a normalization from the definition of $q_\theta$, while symmetry ensures that firms have no reason to favor either technology based on differences between the goods markets.

that were imperfectly substitutable in demand or production, as in the existing literature on directed innovation. Improvement in one technology would then trigger relative price and market size adjustments that affect the profits for the other.[19] I exclude these competition and "demand-pull" forces to focus instead on how the innovation process itself shapes firms' incentives. The mRNA case study provides just one example where these assumptions are realistic, but the forces I study arise even with competition in the goods market.

**Innovation and Firm Dynamics.** Intermediate qualities are determined endogenously through innovation, entry, and exit. Each firm can raise the quality of its intermediates by employing scientists to conduct research. If a firm with qualities $q(t)$ employs $s_\theta(q(t),t)$ scientists to research technology $\theta$ at $t$, the quality of its intermediate evolves according to

$$\dot{q}_\theta(t) = [\lambda q_\theta(t) + \sigma_I K_\theta(t)]\,\eta_\theta s_\theta(q(t),t). \tag{8}$$

Here $\eta_\theta > 0$ denotes the basic productivity of research for $\theta$. This productivity is augmented by the accumulation of knowledge that raises research productivity, represented by the term in brackets. A central feature of my theory is that knowledge accumulates through two channels: First, following an extensive literature on endogenous growth initiated by Romer (1990), I suppose that each firm can learn from innovations made by all others. This public knowledge is technology-specific, and it is embodied in each technology's *knowledge stock* $K_\theta(t) > 0$. The knowledge stock reflects, for example, all information about technology $\theta$ found in the patent data and scientific publications, or shared among scientists at conferences. Its initial value $K_\theta(0) > 0$ is exogenous, and it increases as firms innovate for $\theta$:

$$\dot{K}_\theta(t) = \int \underbrace{[\lambda q_\theta(t) + \sigma_I K_\theta(t)]\,\eta_\theta s_\theta(q(t),t)}_{\dot{q}_\theta(t)} dF(q(t),t). \tag{9}$$

Let $K(t) \equiv (K_A(t), K_B(t))$ denote the vector of knowledge stocks at time $t$. Second, I make the novel assumption that knowledge also accumulates *within each firm*, embodied in its quality $q_\theta(t)$. This internal knowledge is again technology-specific, and it represents any information produced by the firm that disproportionately improves its own future research efforts. The parameters $\lambda, \sigma_I \geq 0$ in (9) control the extent to which firms draw on internal and public knowledge when innovating, respectively.

Several mechanisms could induce knowledge accumulation within firms. For example, intermediates for a technology $\theta$ could represent distinct research lines, and the knowledge

[19] These price and market size effects are detailed in Acemoglu (1998, 2002), and they are applied to study the direction of innovation between clean and dirty technologies in Acemoglu et al. (2012).

generated within a research line may be more useful for future innovation than knowledge generated by others. A firm may work as a mechanism to coordinate scientists on a single research line and accelerate innovation. The mRNA case study reflects this interpretation: mRNA technology can be used to treat or prevent a variety of different diseases, including genetic conditions, cancer, and viral infections like COVID-19. Each application builds on and contributes to our general understanding of mRNA technology. But the technology must also be tailored to each case, and the leading mRNA firms have indeed specialized within particular applications.[20] Alternatively, work dating back to Marshall (1890) documents that innovation is spatially concentrated, suggesting that knowledge spillovers between researchers happen through direct communication.[21] A firm may employ researchers to facilitate communication or ensure physical proximity in a common office or lab, again catalyzing "internal" knowledge spillovers. Both of these mechanisms imply a role for the firm *per se* in the innovation process, rooted in the insights of Coase (1937) and Alchian and Demsetz (1972) that firms form to internalize externalities and coordinate complementary activities. This role for the firm is essential to the model, and I discuss its implications throughout the analysis below.

Consistent with these microfoundations, I assume that scientists are organized into teams managed by entrepreneurs. A firm innovates by employing one entrepreneur and her team to research its intermediates. For simplicity, I assume that each entrepreneur manages the same number $s \equiv S/N$ of scientists, so all innovating firms have equally sized "R&D departments." However, a fraction $1-N$ of firms are unable to innovate at each time because entrepreneurs are in scarce supply. The team size $s$ is a natural measure of the *concentration of scientists* across firms: Holding the total mass of scientists $S$ fixed, an increase in $s$ reduces the number of firms $N$ that can innovate while raising the R&D resources available to any firm that still can. To reduce notation without any essential changes to the model, I assume that any firm unable to hire an entrepreneur must exit. The distribution $F(q,t)$ describes the qualities for active firms and is scaled by their total measure $N > 0$.

In this baseline model, firm entry and exit are exogenous: Each incumbent receives an independent exit shock at rate $\delta > 0$, at which time it ceases production and is replaced by an entrant with initial qualities

$$q_\theta^E(t) \equiv \sigma_E K_\theta(t). \tag{10}$$

Here $\sigma_E \geq 0$ determines the strength of knowledge spillovers to entrants. The entrant immediately employs the entrepreneur and scientists from the exiting firm.

[20] Moderna has primarily focused on developing mRNA vaccines for contageous diseases, while BioNTech has instead pursued mRNA vaccines for cancer.

[21] For example, see Jaffe et al. (1993) and Kalyani et al. (2023).

Before exit, I assume that all profits generated by an innovating firm accrue to its entrepreneur and scientists. The scientists are then allocated at each time to maximize the firm's value $V(q,t)$, taking as given its initial qualities $q$ and the trajectories of the knowledge stocks $[K(t)]_t$ and the interest rate $[r(t)]_t$:

$$V(q,t) = \max_{[s_\theta(q(\tau),\tau)]_{\theta,\tau}} \int_t^\infty \exp\left[-\int_0^\tau (r(t')+\delta)dt'\right]\pi(q(\tau))d\tau, \tag{11}$$

where maximization is subject to the resource constraint $s_A(q,\tau)+s_B(q,\tau) \leq s$ and the quality evolution equation (8). Asset market clearing requires that the total demand for assets from the consumer equal the total value of incumbents at each time:

$$\mathcal{A}(t) = \int V(q,t)dF(q,t). \tag{12}$$

This specification of innovation, entry, and exit implies that the distribution $F(q,t)$ evolves according to the Kolmogorov forward equation (KFE)

$$\frac{\partial F(q,t)}{\partial t} = -\int \dot{q}_A(t)F\left(q_A, dq_B', t\right) - \int \dot{q}_B(t)F\left(dq_A', q_B, t\right) + \delta N \mathbb{1}_{q \geq q^E(t)} - \delta F(q,t). \tag{13}$$

The mass of incumbents with qualities below $q$ declines as incumbents with technology $A$ qualities $q_A$ improve their $A$ intermediates. The first term on the right side of (13) captures the corresponding loss of mass per unit of time, or *flux*, through the boundary $\left\{q' : q_A' = q_A\right\}$. The second term similarly captures the flux through the boundary $\left\{q' : q_B' = q_B\right\}$. The third term reflects the increase in mass from entry, while the last term gives the fall in mass from exit.

**Definition 1.** An *equilibrium* is a set of trajectories for total output $[Y(t)]_t$, consumption $[C(t)]_t$, assets $[\mathcal{A}(t)]_t$, labor demand $[L(t)]_t$, intermediate quantities $[x_\theta(q,t)]_{\theta,q,t}$, wages $[w_L(t)]_t$, intermediate prices $[p_\theta(q,t)]_{\theta,q,t}$, knowledge stocks $[K(t)]_t$, allocations of scientists $[s_\theta(q,t)]_{\theta,q,t}$, incumbent values $[V(q,t)]_{q,t}$, and the quality distribution $[F(q,t)]_{q,t}$ such that

(i) the representative consumer chooses her consumption and asset holdings to maximize her utility (1) subject to her budget constraint (2) and no-Ponzi condition (3);

(ii) labor demand and intermediate quantities are chosen by the final producer to maximize profits, given input prices;

(iii) intermediate prices are chosen to maximize incumbent flow profits (7);

(iv) the markets for labor, goods, and assets clear (5, 6, 12);

(v) the knowledge stocks satisfy the evolution equation (9);

(vi) the incumbent value function satisfies (11), while the scientist allocation solves the corresponding maximization problem; and

(vii) the quality distribution satisfies the KFE (13).

I maintain several parameter restrictions throughout the analysis of equilibrium. I assume $\lambda + \sigma_I > 0$ so that innovation is possible, though either of the parameters $\lambda$ or $\sigma_I$ may equal zero. I also assume that spillovers to entrants are positive $\sigma_E > 0$, which ensures that the economy generates long-run growth. Finally, I assume that the exit rate $\delta$ is sufficiently large that an initial cohort of incumbents cannot generate aggregate growth in the long run:

$$\delta > (\lambda + \sigma_I N) \max\{\eta_A, \eta_B\} s. \tag{14}$$

## 3.2 Equilibrium: Static Block

To simplify the characterization of equilibrium, I note that the quality distribution $F$ and the knowledge stocks $K \equiv (K_A, K_B)$ are the state variables in this economy. With the exception of the incumbent value function $V$ and the allocation of scientists $[s_\theta(q,t)]_{\theta,q,t}$, all remaining equilibrium variables are statically determined as a function of the state $(F, K)$. In fact, only the *aggregate qualities* $Q \equiv (Q_A, Q_B)$ are needed to determine this "static block" of the economy, where the aggregate quality of technology $\theta$ is

$$Q_\theta(t) \equiv \int q_\theta dF(q,t).$$

Integration by parts reveals that the aggregate quality of $\theta$ increases with incumbent innovation but potentially declines as incumbents exit:[22]

$$\dot{Q}_\theta(t) = \dot{K}_\theta(t) + \delta\left[N q_\theta^E(t) - Q_\theta(t)\right]. \tag{15}$$

The following lemma characterizes the static block of the economy along with the equilibrium interest rate $r(t)$:

**Lemma 1.** *In equilibrium,*

*(i) the interest rate is* $r(t) = \rho$*;*

[22]The convention in the endogenous growth literature is to identify the knowledge stock $K_\theta$ with the aggregate quality $Q_\theta$. This literature generally does not consider entrants that may arrive with lower qualities than the average incumbent ($\sigma_E < 1$), which directly reduces aggregate quality. I introduce the distinction between $K_\theta$ and $Q_\theta$ so that public knowledge is not artificially reduced by firm entry and exit.

*(ii) intermediate prices and the wage for workers are*

$$p_\theta(q,t) = \frac{\gamma}{1-\beta},$$
$$w_L(t) = [Q_A(t) + Q_B(t)]\,\frac{\beta \bar{x}^{1-\beta} L^{-(1-\beta)}}{1-\beta};$$

*(iii) production labor demand is $L(t) = L$;*

*(iv) intermediate quantities and the corresponding flow profis are*

$$x_\theta(q,t) = q_\theta \bar{x}, \quad \textit{where} \quad \bar{x} \equiv L\left(\frac{\gamma}{1-\beta}\right)^{-\frac{1}{\beta}},$$
$$\pi(q_\theta) = q_\theta \bar{\pi}, \quad \textit{where} \quad \bar{\pi} \equiv \beta L\left(\frac{\gamma}{1-\beta}\right)^{-\frac{1-\beta}{\beta}};$$

*(v) total output and consumption are*

$$Y(t) = [Q_A(t) + Q_B(t)]\,\bar{Y}, \quad \textit{where} \quad \bar{Y} \equiv \frac{\bar{x}^{1-\beta} L^{\beta}}{1-\beta},$$
$$C(t) = [Q_A(t) + Q_B(t)]\,\bar{C}, \quad \textit{where} \quad \bar{C} \equiv \bar{Y} - \gamma \bar{x}.$$

All proofs are found in Appendix A. The wage $w_L(t)$, total output $Y(t)$, and consumption $C(t)$ are all linear in the sum of the aggregate qualities $Q_A(t) + Q_B(t)$, again reflecting the absence of substitution between technologies $A$ and $B$. As noted above, the fixed interest rate $r(t) = \rho$ aids tractability. In the remaining analysis, I focus on the "dynamic block" of the economy: firm innovation decisions and the resulting dynamics of the state $(F, K)$.

## 3.3 Balanced Growth

The main goal of the analysis is to characterize innovation decisions and transitional dynamics after a "new technology" $B$ is introduced into the economy. To lay a foundation for this exercise, I first study the economy's steady states, or *balanced growth paths* (BGPs). The characterization of BGPs in Proposition 1 yields the first core result of the theory: When firms accumulate knowledge internally ($\lambda > 0$), increasing the concentration of scientists $s$ always raises the economy's BGP growth rate. It also produces changes in the firm-quality distribution $F$ that play a critical role in the analysis of technology choice in Section 4.

Toward a definition of balanced growth, note that the distribution of qualities $F$ is never stationary in equilibrium: As firms innovate and contribute to the knowledge stocks $K_\theta$, they improve the initial qualities for entrants $q_\theta^E(t) = \sigma_E K_\theta(t)$. This process repeats as entrants begin innovating, and it produces an upward shift in the distribution of qualities over time. We can attempt to stabilize the distribution by normalizing the qualities $q_\theta$ by the quality of an entrant $q_\theta^E(t)$. Define the *relative quality* of an intermediate by

$$z_\theta(t) \equiv \frac{q_\theta(t)}{q_\theta^E(t)}, \tag{16}$$

and denote a firm's vector of relative qualities by $z = (z_A, z_B)$. Let $H(z,t) \equiv F(zq^E(t), t)$ denote the corresponding distribution, where $zq^E(t)$ denotes the pointwise product $(z_A q_A^E(t), z_B q_B^E(t))$.

**Definition 2.** A *balanced growth path (BGP)* is an equilibrium in which

(i) $K_\theta$ and $Q_\theta$ grow at a constant rate $g_\theta^* \geq 0$ for each technology $\theta \in \{A, B\}$; and

(ii) the relative quality distribution is stationary: $H(z,t) = H^*(z)$ for all $t \geq 0$.

We can observe immediately that the economy admits at most three BGPs, only two of which are locally stable. For each technology $\theta$, there exists a stable BGP in which all firms exclusively innovate for $\theta$. As a result, technology $\theta$ grows asymptotically while the other technology $\theta'$ permanently stagnates. These two very different BGPs arise because innovation for each technology features dynamic increasing returns: Innovation for $\theta$ produces knowledge externally (through $K_\theta$) and within the firm (through $q_\theta$) that raises the productivity of future research for $\theta$. With positive spillovers to entrants $\sigma_E > 0$, any technology that develops a large enough lead in accumulated knowledge attracts all innovation by entrants, reinforcing its advantage. Note that without substitution between technologies $A$ and $B$ in demand or production, there are no relative price adjustments that could redirect innovation toward the lagging technology and ensure a unique "interior" BGP.

Whenever the two stable BGPs exist, there also exists a third BGP that features equal and positive growth rates for both technologies. All scientists must be evenly split across the technologies to sustain equal growth rates. However, this BGP is fragile to perturbations in which a majority of scientists are temporarily directed to just one of the two technologies, which can push the economy toward one of the stable "corner" BGPs. Finally, we can rule out BGPs in which the growth rates $g_A^*$ and $g_B^*$ are both positive but unequal: Eventually all entrants would choose to research the technology with the faster growth rate, and the growth rate of the other would be driven to zero.

I restrict attention to the economy's two stable BGPs. To characterize them, suppose all firms permanently direct their scientists toward technology $\theta$. The evolution of intermediate qualities is then completely mechanical:

$$\dot{q}_\theta(t) = [\lambda q_\theta(t) + \sigma_I K_\theta(t)]\,\eta_\theta s \quad \text{and} \quad \dot{q}_{\theta'}(t) = 0.$$

With no innovation directed toward $\theta'$, all relative qualities $z_{\theta'}$ remain fixed at one. The BGP is then summarized by the growth rate $g^*_\theta$ and the marginal distribution $H^*_\theta(z_\theta)$ of relative qualities for $\theta$. We can use the evolution equation (9) for the knowledge stock $K_\theta$ to write the growth rate $g^*_\theta$ as a function of the distribution $H^*_\theta$:

$$g^*_\theta = \frac{\dot{K}_\theta}{K_\theta} = \int [\lambda \sigma_E z_\theta + \sigma_I]\,\eta_\theta s dH^*_\theta(z_\theta). \tag{17}$$

The distribution $H^*_\theta$ is determined by the evolution of qualities through the KFE (13). To see this, let $F_\theta(q_\theta, t)$ denote the non-stationary marginal distribution of qualities for technology $\theta$. This distribution satisfies the following one-dimensional version of the KFE:

$$\frac{\partial F_\theta(q_\theta, t))}{\partial t} = -[\lambda q_\theta + \sigma_I K_\theta(t)]\,\eta_\theta s f_\theta(q_\theta, t) + \delta\left[N \mathbb{1}_{q_\theta \geq q^E_\theta(t)} - F_\theta(q_\theta, t)\right].$$

Here $f_\theta(q_\theta, t) \equiv \partial F_\theta(q_\theta, t)/\partial q_\theta$ denotes the density of $F_\theta$. The first term denotes the loss of mass as firms with the quality $q_\theta$ innovate, while the second term denotes the net change in mass due to entry and exit. Since the equation $H^*_\theta(z_\theta) = F_\theta(z_\theta q^E_\theta(t), t)$ must hold for all $t$ along a BGP, we can differentiate to find a time-invariant differential equation for $H^*_\theta$:

$$0 = -\left[(\lambda - g^*_\theta) z_\theta + \frac{\sigma_I}{\sigma_E}\right] h^*_\theta(z_\theta) + \delta N - H^*_\theta(z_\theta). \tag{18}$$

The solution to this differential equation gives the distribution $H^*_\theta$ as a function of the growth rate $g^*_\theta$. Candidates for the BGP growth rate $g^*_\theta$ and stationary distribution $H^*_\theta$ must solve the system (17, 18).

To ensure that a solution exists and delivers finite values for firms and the consumer, I maintain the following parameter restrictions:

$$s > \sigma_E S, \tag{19}$$

$$\rho > -\frac{\delta - \lambda \eta_\theta s - \sigma_I \eta_\theta S}{2} + \sqrt{\left(\frac{\delta - \lambda \eta_\theta s - \sigma_I \eta_\theta S}{2}\right)^2 + (\lambda \sigma_E + \sigma_I)\,\delta \eta_\theta S}. \tag{20}$$

The first condition (19) ensures that spillovers to entrants are sufficiently small that an incumbent's quality $q_\theta(t)$ grows faster than the entrant quality $q^E_\theta(t)$.[23] This assumption implies that relative qualities $z_\theta(t)$ are weakly above one, which is essential to obtain a non-degenerate stationary distribution $H^*_\theta$. The second condition (20) ensures that the discount rate $\rho$ is large enough that the consumer's discounted utility is finite in equilibrium. Together with the lower bound on the exit rate (14), it is also sufficient to ensure that firm values are finite.

Given these assumptions, the first proposition characterizes the economy's stable BGPs:

**Proposition 1.** *The economy has two locally stable BGPs, one for each technology $\theta$. In the BGP for technology $\theta$:*

*(i) All scientists research technology $\theta$.*

*(ii) The knowledge stock $K_\theta$ and aggregate quality $Q_\theta$ grow at rate*

$$g^*_\theta = -\frac{\delta - \lambda\eta_\theta s - \sigma_I \eta_\theta S}{2} + \sqrt{\left(\frac{\delta - \lambda\eta_\theta s - \sigma_I \eta_\theta S}{2}\right)^2 + (\lambda\sigma_E + \sigma_I)\delta\eta_\theta S}.$$

*(iii) The stationary distribution $H^*_\theta$ is a generalized Pareto distribution with location parameter 1, shape parameter $\varphi^*_\theta > 0$, and tail parameter $\xi^*_\theta \in (-\infty, 1)$:*

$$H^*_\theta(z_\theta) = \begin{cases} N\left[1 - \left(1 + \xi^*_\theta \frac{z_\theta - 1}{\varphi^*_\theta}\right)^{-\frac{1}{\xi^*_\theta}}\right] & \xi^*_\theta \neq 0, \\ N\left[1 - \exp\left(-\frac{z_\theta - 1}{\varphi^*_\theta}\right)\right] & \xi^*_\theta = 0. \end{cases}$$

*The parameters satisfy*

$$\varphi^*_\theta = \frac{\left(\lambda + \frac{\sigma_I}{\sigma_E}\right)\eta_\theta s - g^*_\theta}{\delta} \quad \text{and} \quad \xi^*_\theta = \frac{\lambda\eta_\theta s - g^*_\theta}{\delta}.$$

Along the BGP for technology $\theta$, all aggregates including the knowledge stock $K_\theta$, aggregate quality $Q_\theta$, total output $Y$, and consumption $C$ grow at the common rate $g^*_\theta > 0$. This growth rate is naturally increasing in the extent of knowledge spillovers ($\sigma_E$, $\sigma_I$) and the productivity of research ($\lambda$, $\eta_\theta$). With fully endogenous growth, the model also features a scale effect whereby the growth rate is increasing in the total mass of scientists $S$. The growth rate is declining in the exit rate $\delta$ because all innovation is undertaken by incumbents.

The most interesting comparative static for $g^*_\theta$ concerns the concentration of scientists $s$. Provided that incumbents build to some extent on their own past advances ($\lambda > 0$), the growth rate strictly increases with the concentration of scientists. This observation clarifies a role for

[23] A condition is needed because $q_\theta(t)$ grows with innovation by a single firm while $q^E_\theta(t) = \sigma_E K_\theta(t)$ grows with spillovers from all firms.

*the firm* per se in the growth process: If scientists build on the knowledge generated by others, and if these spillovers are particularly intense between scientists at the same firm, then larger firms can better catalyze spillovers and accelerate innovation. Any of the microfoundations discussed in Section 3.1 could account for this "theory of the firm," and Proposition 1 draws a sharp implication for balanced growth.

The second part of Proposition 1 shows that the BGP relative quality distribution falls into the familiar generalized Pareto class. The shape and tail parameters of the distribution depend endogenously on the growth rate $g_\theta^*$ and the parameters of the innovation process. To understand this relationship, note that an intermediate's relative quality $z_\theta(t)$ increases as the growth rate of its quality outpaces the aggregate growth rate $g_\theta^*$:

$$\frac{\dot{z}_\theta(t)}{z_\theta(t)} = \frac{\dot{q}_\theta(t)}{q_\theta(t)} - \frac{\dot{K}_\theta(t)}{K_\theta(t)} = \left[\lambda + \frac{\sigma_I}{\sigma_E} z_\theta(t)^{-1}\right]\eta_\theta s - g_\theta^*.$$

This calculation follows directly from the evolution equation (8) for $q_\theta(t)$. Proposition 1 shows that the numerator of the shape parameter $\varphi_\theta^*$ is equal to the growth rate of relative quality for an entrant ($z_\theta(t) = 1$). The numerator of the tail parameter $\xi_\theta^*$ is instead the limiting growth rate for a long-lived incumbent ($z_\theta(t) \to \infty$). An increase in either parameter reallocates mass toward larger relative qualities; the shape parameter primarily affects the "body" of the distribution (low $z_\theta$), while the tail parameter naturally affects the tail (high $z_\theta$). Hence any change in the model primitives that raises the growth rate of a firm relative to the growth rate of the economy will produce a more skewed distribution. Because it plays a key role in the analysis of technology choice in Section 4, the following corollary formally shows that this holds for the concentration of scientists $s$:

**Corollary 1.** *The shape parameter $\varphi_\theta^*$ is strictly increasing in $s$. The tail parameter $\xi_\theta^*$ is strictly increasing in $s$ if and only if $\lambda > 0$.*

This result is not immediate, because an increase in $s$ raises both the growth rate of a firm $\frac{\dot{q}_\theta(t)}{q_\theta(t)}$ and the economy's growth rate $g_\theta^*$ when $\lambda > 0$. The first effect is larger because aggregate growth is generated by firm quality growth *in combination with* entry and exit. The latter attenuates the link between aggregate growth and firm quality growth. I prove additional comparative statics for $\varphi_\theta^*$ and $\xi_\theta^*$ in Appendix A.

# 4 Equilibrium with Technology Choice

In this section, I analyze the economy's equilibrium when firms can choose to innovate for technology $A$ or technology $B$. I suppose that initially only $A$ is available, and that incumbents have innovated for $A$ before $B$ arrives at $t = 0$. These technologies pose a simple trade-off: $B$ has the higher basic research productivity $\eta_B > \eta_A$, so it can support faster growth in the long run. But its initial knowledge stock $K_B(0) > 0$ may be lower than that for $A$, limiting growth in the short run. Initial incumbents at $t = 0$ are endowed with the entrant quality $q_B^E(0)$, so that no firm has any absolute advantage in innovating $B$ to start. To study the effect of market structure on the aggregate technology choice, I suppose an arbitrary initial distribution $H_{A0}(z_A)$ of relative qualities for $A$. I later endogenize it using the BGP distribution $H_A^*$ from Proposition 1 to explore how the concentration of scientists $s$ affects the equilibrium.

I begin by characterizing the solution to the firm's problem (11) in Section 4.1. Firm innovation decisions are generally *path-dependent* and *forward-looking*, and knowledge spillovers generate two kinds of strategic complementarities that critically shape (and complicate) these decisions. To resolve the resulting technical challenges, I study the class of *monotone equilibria* in Section 4.2. These equilibria exist under weak conditions, and they display the pattern of innovation observed in the mRNA case study (Section 2) and the empirical analysis (Section 5): Initial incumbents are reluctant to innovate for the new technology $B$, but they may begin doing so after entrants have made substantial progress. The initial market structure $H_{A0}$ plays a critical role, as incumbents with greater expertise for $A$ are more reluctant to innovate for $B$. In Section 4.3 I discuss the benchmark case with no knowledge spillovers across incumbents ($\sigma_I = 0$), which features a unique, closed-form equilibrium. When the economy is initially following the BGP for $A$ ($H_{A0} = H_A^*$), I show that the concentration of scientists $s$ has an ambiguous effect on the equilibrium direction of innovation. But it tends to slow or prevent a transition to $B$ provided the discount rate $\rho$ is sufficiently high. The social planner may not always transition to $B$ *more often* than in equilibrium, but when it is optimal, the social planner always transitions *more quickly*.

## 4.1 The Firm's Problem

Consider a firm with qualities $q$ at time $t \geq 0$. Given a trajectory for the knowledge stocks $[K(\tau)]_\tau$, the firm's problem (11) is to allocate its scientists across its intermediates at all times $\tau \geq t$ to maximize the present value of its profits. To understand the underlying incentives, note that the firm's marginal value of research (scientists) for technology $\theta$ at $t$ satisfies

$$\frac{dV(q,t)}{ds_\theta(q,t)} = \frac{\partial V(q,t)}{\partial q_\theta} [\lambda q_\theta + K_\theta(t)] \eta_\theta. \tag{21}$$

From the evolution equation (8) for $q_\theta(t)$, an additional scientist at $t$ raises the quality $q_\theta$ by $[\lambda q_\theta + K_\theta(t)] \eta_\theta$, increasing the firm's value in proportion to $\partial V(q,t)/\partial q_\theta$. The marginal value of research $dV(q,t)/ds_\theta(q,t)$ plays a key role in the firm's problem, because any solution must allocate scientists only to the technology with the larger marginal value at each $t$. The next proposition characterizes this marginal value by showing how the firm's value $V(q,t)$ varies with its qualities $q$. Together with (21), this result clarifies the path-dependent and forward-looking forces as well as the impact of spillovers on the firm's innovation decisions.

**Proposition 2.** *The incumbent's marginal value of quality for technology $\theta$ is*

$$\frac{\partial V(q,t)}{\partial q_\theta} = \bar{\pi}\Psi_\theta(t),$$

$$\text{where} \qquad \Psi_\theta(t) \equiv \int_t^\infty \exp\left(-\int_t^{t'} [\rho+\delta-\lambda\eta_\theta s_\theta(q(\tau),\tau)]\,d\tau\right) dt'.$$

*The marginal value of research for $\theta$ is then*

$$\frac{dV(q,t)}{ds_\theta(q,t)} = \bar{\pi}\Psi_\theta(t)[\lambda q_\theta + \sigma_I K_\theta(t)]\,\eta_\theta.$$

*When $\lambda > 0$, this marginal value is strictly increasing in the current quality $q_\theta$ and the mass of scientists $s_\theta(q(\tau),\tau)$ researching $\theta$ at future times $\tau > t$. When $\sigma_I > 0$, the marginal value is strictly increasing in the current knowledge stock $K_\theta(t)$. It does not depend on technology $\theta' \neq \theta$.*

The expression for the marginal value of quality $\partial V(q,t)/\partial q_\theta$ follows by integrating the evolution equation (8) for $q_\theta(t)$ and substituting into the firm's objective (11). Given an increase in the quality $q_\theta$, the function $\Psi_\theta(t)$ aggregates the additional profits at each time $t' \geq t$, discounted back to $t$. Note that the discount rate at $\tau \geq t$ is $\rho+\delta-\lambda\eta_\theta s_\theta(q(\tau),\tau)$, where the last term reflects the additional profits generated because the innovation raises the productivity of future research for technology $\theta$. The comparative statics for the marginal value of research $dV(q,t)/ds_\theta(q,t)$ in Proposition 2 are immediate from the corresponding expression.

Proposition 2 yields the second core result of the theory: When knowledge accumulates within firms ($\lambda > 0$), innovation decisions are both *path-dependent* and *forward-looking* at the firm level. Path dependence arises because a firm's past innovation for technology $\theta$ raises its current research productivity for $\theta$, incentivizing continued innovation in that direction. This is immediate from Proposition 2, because the marginal value of research $dV(q,t)/ds_\theta(q,t)$ is strictly increasing in the current quality $q_\theta$ when $\lambda > 0$. But history is not always destiny, because the firm also takes into account how its current innovation affects its future research productivities. This is also apparent from Proposition 2, because with $\lambda > 0$ the marginal value of research $dV(q,t)/ds_\theta(q,t)$ increases with future research for $\theta$ via the aggregation factor $\Psi_\theta(t)$. Holding fixed any effect of spillovers on the firm's direction, this forward-looking force pushes the firm to innovate for $B$: The future allocation of scientists is also chosen by the firm, and $B$ allows for faster quality *growth* through cumulative innovation ($\eta_B > \eta_A$). A key insight from the model is that the path-dependent and forward-looking forces arise from the same assumptions that knowledge is technology-specific and accumulates within the firm.

With stochastic entry and exit, these features of the innovation process naturally generate heterogeneity in innovation incentives across firms. Through their past innovation, initial incumbents have higher initial equalities $q_A(0)$ for technology $A$ than entrants, but they share the same initial qualities $q_B^E(0)$ for technology $B$. This gives initial incumbents an endogenous comparative advantage *in innovating A*, incentivizing them to continue with $A$ after $B$ arrives. In this way, the theory readily explains the initial innovation decisions made by firms in the mRNA case study: Entrants opted to explore the potentially promising mRNA technology, while incumbent pharmaceutical firms with existing expertise "opted to invest [their] resources elsewhere" (Dolgin, 2021).

Knowledge accumulation within firms is crucial for these results. When firms instead draw only on aggregate knowledge ($\lambda = 0$), the marginal value of research simplifies to

$$\frac{dV(q,t)}{ds_\theta(q,t)} = \frac{\bar{\pi}}{\rho + \delta}\sigma_I \eta_\theta K_\theta(t).$$

Each firm allocates its scientists to the technology with the larger marginal value at each $t$, so in this case firm heterogeneity has no impact on innovation decisions. The assumption of purely public knowledge accumulation is pervasive in the endogenous growth and directed innovation literatures, precluding any meaningful role for the firm (or the concentration of scientists $s$) in the growth process. See Appendix C for a detailed discussion of this case.

Proposition 2 is also suggestive of the third core result of the model, developed in greater detail in Sections 4.2 and 4.3: A firm's innovation decisions are influenced by spillovers from other firms, providing a mechanism by which the decisions made by initial incumbents can persistently affect the aggregate direction of innovation. The model features two sources of spillovers. First, the initial qualities of an entrant are determined by the aggregate knowledge stocks, $q_\theta^E(t) = \sigma_E K_\theta(t)$. When firm innovation decisions are path-dependent ($\lambda > 0$), past innovation for technology $\theta$ then raises the entrant's marginal value of research for $\theta$. These "backward-looking" spillovers generate a strategic complementarity between past and current firms. A second source of spillovers arises when $\sigma_I > 0$, so that incumbents draw on public knowledge when innovating. In this case, a firm has incentives to innovate for the technology with the larger current knowledge stock $K_\theta(t)$, generating a strategic complementarity between past and current firms and across all current firms. If in addition $\lambda > 0$, so that the firm's choice of direction is forward-looking, then the firm must account for the future path of the knowledge stocks: The expectation of high growth in $K_\theta(t)$ may induce the firm to innovate for technology $\theta$, even if its current research productivity is lower.

The analysis above provides insight into firm innovation incentives, but it does not yield a sharp characterization of innovation decisions. Two difficulties remain. First, when $\lambda > 0$, the firm's marginal value of research for a given technology $\theta$ is increasing in its past and future research for $\theta$. These complementarities render the firm's problem non-convex, so first-order conditions are not sufficient to characterize the solution.[24] Second, knowledge spillovers across incumbents can complicate innovation decisions and introduce multiple equilibria. For example, an incumbent innovating for $\theta$ may switch to $\theta'$ if the other knowledge stock $K_{\theta'}(t)$ is growing (or expected to grow) quickly enough. Moreover, with knowledge spillovers across incumbents ($\sigma_I > 0$) and knowledge accumulation within incumbents ($\lambda > 0$), incumbents are more likely to innovate for a technology $\theta$ if they expect its knowledge stock $K_\theta(t)$ to grow in the future. Since the knowledge stock grows more rapidly as more firms innovate for $\theta$, these spillovers can produce multiple equilibria.

To address the first difficulty, I reformulate the firm's problem (11) as an optimal stopping problem. As noted above, at each time the firm optimally allocates all scientists to the technology $\theta$ with the larger marginal value of research. As a result, I can cast the firm's problem as the choice of an initial innovation direction $\theta_0 \in \{A, B\}$ and a sequence of stopping times $t < T_1 \leq T_2 \leq \ldots$ at which the firm completely reverses its direction. The next lemma records this observation and provides a first-order necessary condition – the *smooth-pasting condition* – for interior stopping times.[25]

**Lemma 2.** *Given initial qualities $q$ and a trajectory for the knowledge stocks $[K(\tau)]_\tau$, for any solution to the firm's problem (11) there exists an initial innovation direction $\theta_0 \in \{A, B\}$ and a sequence of stopping times $t < T_1 \leq T_2 \leq \ldots$ such that the firm exclusively innovates for $\theta_0$ at $t' \in [t, T_1)$ and reverses its innovation direction at each stopping time $T_k$. Every interior stopping time $T_k \in (t, \infty)$ must satisfy the smooth-pasting condition*

$$\Psi_B(T_k)\left[\lambda q_B(T_k) + \sigma_I K_B(T_k)\right]\eta_B = \Psi_A(T_k)\left[\lambda q_A(T_k) + \sigma_I K_A(T_k)\right]\eta_A. \tag{22}$$

The smooth-pasting condition (22) simply requires that the marginal value of research must be equalized across technologies at each stopping time $T_k$. Lemma 2 provides a much more tractable description of the firm's problem than the infinite-dimensional optimal control problem (11). But it does not address the second difficulty discussed above: With spillovers

[24] For example, it can be optimal for a firm to innovate for $A$ at $t$ assuming it exclusively innovates for $A$ in the future, but it might raise value even further to switch all current and future innovation to $B$.

[25] I adopt the terminology of Dixit (1993) for this optimality condition, though I do not emphasize its connection to the differentiability of the firm's value function at the stopping times.

across incumbents $\sigma_I > 0$, the firm may reverse its research direction repeatedly, and the economy may feature multiple equilibria. In the next section, I define and characterize *monotone equilibria* that rule out repeated reversals of a firm's innovation direction over time.

## 4.2 Monotone Equilibria

When innovation decisions are both inherently forward-looking ($\lambda > 0$) and subject to knowledge spillovers across incumbents ($\sigma_I > 0$), the model features a dynamic strategic complementarity that renders a full characterization of all equilibria intractable. To simplify while maintaining the key economics, I make two restrictions. First, I consider only equilibria that converge to one of the economy's stable BGPs. This restriction excludes cyclical equilibria as well as equilibria that happen to converge to the economy's unstable BGP. Second, I consider only equilibria in which innovation decisions are appropriately monotone over time:

**Definition 3.** An equilibrium is *monotone* if the equilibrium allocation of scientists $s_\theta(q(t), t)$ is monotone in $t \geq 0$ for every possible starting time $t_0 \geq 0$ for a firm. Here $q(t)$ evolves according to (8) for $t > t_0$. If $t_0 > 0$, then $q(t) = q^E(t)$ for $t \leq t_0$.

To understand the definition, first set $t_0 = 0$. Monotonicity then implies that an initial incumbent reverses its innovation direction at most once. By Lemma 2, this reduces the incumbent's problem to a choice of an initial research direction and a *single* stopping time $T$. Instead setting $t_0 > 0$, monotonicity implies the same restriction on entrants, with one additional requirement: If at any time $t > 0$ the initial innovation direction for entrants reverses from $\theta$ to $\theta'$, all subsequent entrants permanently innovate for $\theta'$. In this sense, monotonicity requires that innovation is monotone both within firms and across entrants over time.

I focus on monotone equilibria to maintain analytical tractability while allowing for knowledge spillovers across incumbents ($\sigma_I > 0$), which can generate realistic innovation dynamics. For example, we will see in Section 4.3 that in the absence of these spillovers firms never reverse their initial innovation directions. Spillovers to entrants still generate linkages across firms, but they cannot rationalize why, for example, incumbent pharmaceutical firms are now researching and developing new mRNA therapies. By contrast, spillovers across incumbents can induce a firm to alter its direction of innovation. Monotonicity simply ensures that this occurs at most once for each firm, and so works as a joint restriction on $\sigma_I$ and the trajectories of the knowledge stocks $[K(t)]_t$. The empirical analysis in Section 5 also suggests that these spillovers are important to explain firm patenting behavior, and it shows that incumbent innovation for many new technologies is generally monotone in the sense of Definition 3.

To understand the structure of monotone equilibria, consider one that converges to the BGP for technology $B$. Each initial incumbent chooses an innovation direction $\theta_0$ and a stopping time $T \in (0, \infty]$ at which to *permanently* reverse it. Since the equilibrium converges to the BGP for $B$, it can be shown that any reversal must be from $A$ to $B$. An initial incumbent then either permanently innovates for $B$ or begins innovating for $A$ before switching permanently to $B$ (if at all). These firms differ only by their initial relative qualities $z_A(0)$ for $A$, so path dependence implies a cutoff value $z_{A0} \geq 1$ such that an incumbent starts innovating for $A$ if and only if $z_A(0) > z_{A0}$.

For entrants, there exists a time $T_E \in [0, \infty]$ after which all entrants innovate permanently for one of the technologies; since the equilibrium converges to the BGP for $B$, this must be technology $B$. Any entrant that arrives before $T_E$ then begins by innovating for $A$, but it also chooses a stopping time $T \in (T_E, \infty]$ at which to permanently switch to $B$. To simplify, I note that an equilibrium with $T_E = 0$ exists whenever an equilibrium with $T_E > 0$ exists — the latter is essentially a "translation" of a monotone equilibrium with $T_E = 0$, with all firms innovating for $A$ until $T_E > 0$. I restrict to $T_E = 0$ in what follows.

To characterize the stopping times $T$, note that each such time must satisfy the smooth-pasting condition (22). For a more convenient statement of this condition, denote the total knowledge about technology $\theta$ available to a firm with quality $q_\theta(t)$ by

$$k_\theta(t) \equiv [\lambda q_\theta(t) + \sigma_I K_\theta(t)].$$

In this case, the smooth-pasting condition (22) can then be written

$$\frac{k_B(T)\eta_B}{\rho + \delta - \lambda \eta_B s} = \frac{k_A(T)\eta_A}{\rho + \delta}. \tag{23}$$

Intuitively, at the stopping time $T$ the value of the firm's initial research for $B$ must be equal to the value of its final research for $A$. Given the non-convexities in the firm's problem described in Section 4.1, the smooth-pasting condition (23) alone does not fully characterize a firm's stopping time $T$. But including the associated second-order condition, that the left side of (23) be weakly increasing relative to the right side just before $T$, does yield a full characterization.

Thus a monotone equilibrium is determined by the initial cutoff $z_{A0}$, with all stopping times $T$ characterized by the smooth-pasting condition (23) and the associated second-order condition. The next proposition summarizes this description of a monotone equilibrium converging to $B$. It additionally demonstrates that innovation decisions can be written in terms of a *cutoff function* $\chi(t)$ for the *firm knowledge ratio* $\frac{k_B(t)}{k_A(t)}$. The initial value $\chi(0)$ is related to the relative quality cutoff $z_{A0}$ by the identity

$$\chi(0) = \frac{\lambda \sigma_E + \sigma_I}{\lambda \sigma_E z_{A0} + \sigma_I} \kappa(0),$$

where I define the *aggregate knowledge ratio*

$$\kappa(t) \equiv \frac{K_B(t)}{K_A(t)}. \tag{24}$$

This ratio measures the "gap" between the two technologies, and it plays a key role in the analysis below. The proposition also provides a monotone comparative static for the cutoff function $\chi(t)$ with respect to the initial relative quality distribution $H_{A0}$. I state the analogous result for monotone equilibria converging to the BGP for technology $A$ in Appendix B.

**Proposition 3.** *In any monotone equilibrium converging to the BGP for technology B, there exists a cutoff function $\chi(t)$ such that a firm innovates for B if and only if*

$$\frac{k_B(t)}{k_A(t)} > \chi(t).$$

*The cutoff function satisfies the differential equation (B5), and the monotone equilibrium is unique up to the initial value $\chi(0)$. The knowledge stocks $K(t)$ solve the dynamical system (B10).*

*The initial value $\chi(0)$ is increasing in $H_{A0}$ in the sense of first-order stochastic dominance, and strictly so whenever $\chi(0) \in (\kappa(0), \infty)$ and $\sigma_I > 0$.*

Any monotone equilibrium can be characterized by a simple cutoff function $\chi(t)$ for the firm knowledge ratio $\frac{k_B(t)}{k_A(t)}$, reflecting the dynamic sorting of firms to technologies based on comparative advantage in innovation. Determining this cutoff is not straightforward because firms do not make myopic innovation decisions based only on their current knowledge ratios $\frac{k_B(t)}{k_A(t)}$. The cutoff $\chi(t)$ decreases with the knowledge ratio $\frac{k_B(t)}{k_A(t)}$ of the least-experienced incumbent to innovate for $A$. It remains constant when it reaches the level identified by the smooth pasting condition (22) as the value of the knowledge ratio $\frac{k_B(t)}{k_A(t)}$ at which a firm innovating for $A$ reverses to $B$.

Proposition 3 also demonstrates that both the cutoff $\chi(t)$ and the knowledge stocks $K(t)$ can be recovered by integrating a second-order dynamical system, dramatically simplifying the computation of equilibria. Moreover, monotone equilibria are unique up to the initial condition $\chi(0)$. The key substantive implication of the Proposition concerns the comparative static with respect to $H_{A0}$: The entire cutoff function $\chi(0)$ is increasing in the initial relative quality distribution $H_{A0}$, so that greater initial experience for technology $A$ always slows the transition to technology $B$. This provides a first indication of how the economy's initial market structure $H_{A0}$ can persistently affect aggregate innovation.

As the discussion before Proposition 3 clarifies, monotonicity disciplines equilibrium by ensuring that firms do not reverse their research directions more than once. This may not hold for an arbitrary equilibrium converging to the BGP for technology $B$, because the knowledge stock $K_B(t)$ for technology $B$ may initially decline relative to the knowledge stock $K_A(t)$ for technology $A$. With strong enough spillovers across incumbents, these dynamics could induce an incumbent to start innovating for $B$, reverse to $A$ as $\kappa(t)$ falls, and reverse back to $B$ as $\kappa(t)$ diverges asymptotically. To rule out this effect, in Appendix B I derive a condition on primitives and the trajectory of the knowledge stocks $[K(t)]_t$ to ensure that any firm innovating for $B$ can never satisfy the optimality conditions necessary for a reversal to $A$. The following proposition provides a weaker sufficient condition, exclusively involving primitives, under which monotone equilibria converging to each technology's BGP exist.

**Proposition 4.** *Suppose*

$$\lambda \eta_B s \geq \frac{\sigma_I}{\lambda \sigma_E + \sigma_I} g_A^*.$$

*Then there exist thresholds $\kappa_B^* \leq \kappa_A^*$ such that:*

*(i) A monotone equilibrium converging to $B$ exists if and only if $\kappa(0) \geq \kappa_B^*$.*

*(ii) A monotone equilibrium converging to $A$ exists if and only if $\kappa(0) \leq \kappa_A^*$.*

*The thresholds satisfy $\kappa_B^* < \kappa_A^*$ if and only if $\lambda > 0$ and $\sigma_I > 0$. When $\lambda > 0$, the thresholds are strictly increasing in $H_{A0}$.*

This result shows that an equilibrium converging to the BGP for technology $\theta$ can only exist provided that the initial advantage for $\theta'$ is not too large. Moreover, multiple equilibria can arise for intermediate values of the aggregate knowledge ratio $\kappa(t)$ when the firm's problem (11) is dynamic ($\lambda > 0$) and there are knowledge spillovers across incumbents ($\sigma_I > 0$). As the discussion above indicates, these multiple equilibria are supported by different expectations about other firms' future innovation decisions. Finally, the initial relative quality distribution $H_{A0}$ can have a permanent impact on the economy's aggregate technology choice: Greater initial experience makes it easier to sustain an equilibrium converging to $A$ but harder to sustain an equilibrium converging to $B$. This yields the third core result of the model, demonstrating that an industry's initial market structure can be decisive for its long-run technology choice.

## 4.3 Benchmark: $\sigma_I = 0$

To develop additional intuition, I consider the benchmark case with no spillovers across incumbents ($\sigma_I = 0$). This case provides a simple illustration of equilibrium dynamics, and it provides several new insights about how market structure and the concentration of R&D affect the aggregate direction of innovation. I sketch the equilibrium derivations below, describing firm innovation decisions, their aggregation, comparative dynamics for the concentration of scientists $s$, and potential inefficiencies. I provide full details in Appendix C.

**Firm Innovation.** The case with $\sigma_I = 0$ is particularly tractable because all spillovers between firms are "backward-looking," so each firm is unaffected by other firms' current and future innovation decisions. This conveniently implies that each firm chooses its initial innovation direction with no reversals.

To characterize innovation decisions, let $V^\theta(q,t)$ denote the value from innovating permanently for technology $\theta$ given initial qualities $q$ at time $t$. Integrating the quality evolution equation (8) yields the explicit formula

$$V^\theta(q,t) = \frac{q_\theta + q_{\theta'}}{\rho+\delta} + \frac{1}{\rho+\delta}\frac{\lambda\eta_\theta s}{\rho+\delta-\lambda\eta_\theta s}q_\theta.$$

The first term is the value of the firm's initial qualities $q$, while the second term is the value of subsequent innovation for $\theta$. The firm innovates for $B$ instead of $A$ if an only if $B$ yields a higher value, $V^B(q,t) \geq V^A(q,t)$. Substituting the expression above, this holds if and only if

$$\frac{\eta_B}{\rho+\delta-\lambda\eta_B s}q_B \geq \frac{\eta_A}{\rho+\delta-\lambda\eta_A s}q_A. \tag{25}$$

We can apply this observation to conveniently characterize entrant and incumbent innovation decisions. First consider an entrant at $t \geq 0$. Substituting the entrant qualities $q_\theta^E(t) = \sigma_E K_\theta(t)$ into the inequality (25), we can rearrange to find that it innovates for technology $B$ when the aggregate knowledge ratio $\kappa(t)$ is above the *entry threshold*

$$\kappa^E \equiv \frac{\eta_A}{\rho+\delta-\lambda\eta_A s}\left(\frac{\eta_B}{\rho+\delta-\lambda\eta_B s}\right)^{-1} \in (0,1).$$

Now consider an initial incumbent at $t = 0$ with qualities $q_A(0) = z_A(0)q_A^E(0)$ and $q_B(0) = q_B^E(0)$. Substituting these qualities into (25), we similarly find that the incumbent permanently innovates for technology $B$ when its relative quality $z_A(0)$ is below the cutoff

$$z_{A0}^* \equiv \frac{\kappa(0)}{\kappa^E}$$

This cutoff is inversely related to the initial value of the cutoff function $\chi(t)$ described by Proposition 3, which in this case satisfies $\chi(0) = \kappa^E$. That is, the equilibrium in this case is monotone, and with $\chi(0)$ pinned down it must be unique. Since firms never reverse their innovation directions in equilibrium, there is no need to track the full path of the cutoff function

$\chi(t)$, and I can focus on $z^*_{A0}$ in the analysis below.

**Aggregation.** Using this characterization of innovation decisions, we can directly derive the dynamical system described by Proposition 3 for the evolution of the knowledge stocks $K$. Suppose $\kappa(0) > \kappa^E$, and consider technology $A$. For $t$ near zero, all innovation for $A$ is conducted by initial incumbents that chose not to innovate for $B$. The quality $q_A$ for any such incumbent grows at rate $\lambda\eta_A s$, but these incumbents exit at rate $\delta$. Differentiating the evolution equation (9) for $K_A$ then yields

$$\ddot{K}_A(t) = -(\delta - \lambda\eta_A s)\dot{K}_A(t). \tag{26}$$

With the assumed lower bound on the exit rate (14), this equation implies that the knowledge stock $K_A$ increases more slowly over time as initial incumbents exit and entrants instead choose to innovate for technology $B$. The initial conditions for this differential equation are the initial knowledge stock $K_A(0)$ and the initial growth rate $\dot{K}_A(0)/K_A(0)$, which is decreasing in the cutoff $z^*_{A0}$:

$$\frac{\dot{K}_A(0)}{K_A(0)} = \lambda\eta_A s\sigma_E \int_{z^*_{A0}}^{\infty} z_A dH_{A0}(z_A). \tag{27}$$

Now consider technology $B$. For $t$ near zero, entrants as well as initial incumbents with $z_A(0) \leq z^*_{A0}$ innovate for technology $B$. The quality $q_B$ for any such firm grows at rate $\lambda\eta_B s$, but firms exit at rate $\delta$. Exiting firms are replaced by a mass $\delta N$ of entrants with initial quality $\sigma_E K_B(t)$, so differentiating the evolution equation (9) for $K_B$ yields

$$\ddot{K}_B(t) = -(\delta - \lambda\eta_B s)\dot{K}_B(t) + \lambda\eta_B s\delta N\sigma_E K_B(t). \tag{28}$$

The initial conditions for this differential equation are the initial knowledge stock $K_B(0)$ and the initial growth rate $\dot{K}_B(0)$, which is increasing in the cutoff $z^*_{A0}$:

$$\frac{\dot{K}_B(0)}{K_B(0)} = \lambda\eta_B s\sigma_E H_{A0}(z^*_{A0}). \tag{29}$$

The equations (26, 28) form a second-order, autonomous, linear system of differential equations for the evolution of the key state variables $K$ (while entrants innovate for $B$). This is a natural kind of dynamical system to describe the evolution of an economy undergoing an endogenous technological transition: As Jones (1995) discusses in detail, all models of endogenous *balanced* growth rely on a *first*-order, autonomous, linear differential equation of the form $\dot{X}(t) = gX(t)$ to generate exponential growth, where $X$ is an appropriately-defined knowledge stock or productivity variable.[26] But along an endogenous technological *transition*, the growth rates of the old and new technologies must adjust to accommodate the rise of the new technology. A *second*-order linear system is perhaps the simplest way to describe the evolution of state variables with smoothly changing growth rates.

I show in Appendix C that the system (26, 28) can be integrated in closed form. The solution has two key implications: The aggregate knowledge ratio $\kappa(t)$ depends on initial conditions only through its initial value $\kappa(0)$, in which it is strictly increasing; and $\kappa(t)$ is "U-shaped" over time.[27] These properties are useful because the system (26, 28) only describes the dynamics of the knowledge stocks $K(t)$ while entrants continue to research technology $B$, $\kappa(t) \geq \kappa^E$. If this condition is ever violated, the economy fails to transition in aggregate to technology $B$ and instead converges back to the BGP for technology $A$. Thus there exists a threshold $\kappa^*$ such that the economy converges to the BGP for technology $B$ if and only if $\kappa(0) \geq \kappa^*$. Note that this threshold characterization of the aggregate direction of innovation reflects Proposition 4 for the case in which the equilibrium is unique.

**Equilibrium Transition.** The threshold $\kappa^*$ determines the economy's propensity to transition to technology $B$ in equilibrium, and it depends richly on model primitives. To gain intuition for the key forces, let $H_{A0} = H_A^*$, and note that the analysis above implies that a simple sufficient condition for the economy to transition is for the initial growth rate of $K_B$ to dominate the initial growth rate of $K_A$. By (27, 29), this holds if and only if

$$\eta_B H_A^*\left(z_{A0}^*\right) \geq \eta_A \int_{z_{A0}^*}^{\infty} z_A dH_A^*(z_A). \tag{30}$$

This inequality depends on the initial aggregate knowledge ratio $\kappa(0)$ only through the cutoff $z_{A0}^*$, which is strictly increasing in $\kappa(0)$. As $\kappa(0)$ rises, the left side increases as a larger mass of initial incumbents innovate for $B$, raising the initial growth rate of $K_B$. The right side instead decreases as fewer initial incumbents innovate for $A$, lowering the initial growth rate of $K_A$. There exists a unique value $\bar{\kappa}$ at which the inequality (30) binds:

$$1 = \frac{\eta_A}{\eta_B} \frac{1}{1-\xi_A^*} \frac{\frac{\bar{\kappa}}{\kappa^E}}{\left(\frac{\bar{\kappa}}{\kappa^E}\right)^{1/\xi_A^*} - 1}. \tag{31}$$

[26]For example, the present model satisfies this relation for $X = K_\theta$ along each stable BGP.

[27]The latter holds because exit and entry are sluggish, so that technology $A$ can continue improving relative to technology $B$ until enough entrants have begun innovating for $B$.

Here I have substituted $z^*_{A0} = \frac{\kappa(0)}{\kappa^E}$ and the BGP relative quality distribution $H^*_A$ from Proposition 1, which is a standard Pareto distribution when $\sigma_I = 0$.

The right side of equation (31) is strictly increasing in $\xi^*_A$ and $\kappa^E$ and strictly decreasing in $\tilde{\kappa}$, and it delivers a simple but powerful intuition about the drivers of a technological transition: Any change that thickens the tail of the old technology's firm-quality distribution slows the transition, because it raises both the relative mass of incumbents who choose not to transition and their initial innovation rates. Both effects increase their collective influence over the aggregate direction of innovation, which may be decisive if it induces entrants to switch back to innovating for technology $A$. However, any change that raises incentives for new firms to innovate for the new technology instead accelerates the transition by raising the relative mass of incumbents who choose to transition. The tail parameter $\xi^*_A$ and the entry threshold $\kappa^E$ respectively capture these "composition" and "growth" forces, but they depend on many of the same model primitives. The following proposition provides an explicit comparative static for $\bar{\kappa}$ with respect to the concentration of scientists $s$:

**Proposition 5.** *There exists a discount rate $\bar{\rho} \geq 0$ that depends on model primitives such that $\bar{\kappa}$ is strictly increasing (decreasing) in $s$ locally if and only if $\rho$ is larger (smaller) than $\bar{\rho}$.*

The trade-off between the composition and growth effects hinges on the discount rate $\rho$. When $\rho$ is relatively high, firms are relatively myopic, and the increase in the growth rate of technology $B$ with $s$ has little effect on firm incentives. The composition effect then dominates, and the increase in the concentration of scientists can delay or prevent a transition to technology $B$. In Section 6, I provide a simple calibrated example to show that an increase in the concentration of scientists can raise the thresholds $\bar{\kappa}$ and $\kappa^*$ for reasonable values of $\rho$.

To show that this may be inefficient, I suppose that a social planner instead chooses the initial cutoff $z_{A0}$ to maximize the representative consumer's utility (1) given the initial knowledge stocks $K(0)$, with all innovation decisions by entrants as in equilibrium. The following proposition characterizes properties of the solution $\hat{z}_{A0}$:

**Proposition 6.** *A solution $\hat{z}_{A0}$ to the social planner's problem exists and depends on $K(0)$ only through $\kappa(0)$. There exists a threshold $\hat{\kappa}$ such that*

*(i) the solution $\hat{z}_{A0}$ yields a transition to technology $B$ if and only if $\kappa(0) \geq \hat{\kappa}$;*

*(ii) $\hat{z}_{A0} > z^*_{A0}$ if $\kappa(0) \geq \hat{\kappa}$; and*

*(iii) $\hat{z}_{A0} \leq z^*_{A0}$ if $\kappa(0) < \hat{\kappa}$, with equality only if $z^*_{A0} = 1$.*

In general, the transition thresholds for the social planner $\hat{\kappa}$ and the equilibrium $\kappa^*$ cannot be ranked. This holds because the social planner internalizes knowledge spillovers on future entrants when choosing the long-run direction of innovation, but these spillovers are not necessarily always larger for a given technology: Technology $B$ spillovers are larger in the long-run given $\eta_B > \eta_A$, but technology $A$ spillovers may be larger in the short-run given incumbents' initial expertise for technology $A$ (i.e., the initial distribution $H_A^*$). However, Proposition 6 shows that for a given long-run innovation direction, the social planner always prefers to direct greater initial innovation in that direction than in equilibrium.

# 5 Empirical Analysis

In this section, I present an empirical analysis of firm innovation decisions using patent publications. I document three facts consistent with the theory: First, a firm's current patenting is highly correlated with its previous patenting, controlling for other determinants of innovation outcomes (Section 5.2). Second, for the collection of new technologies identified by Kalyani et al. (2023), a firm's current patenting for a given technology is better predicted by past patenting *within that technology* than patenting in general (Section 5.3). Third, incumbents with greater patenting experience patent *less* for a new technology than less-experienced firms, though this gap shrinks as the technology matures (Section 5.4). The first two facts support the theory's key assumptions that knowledge is both cumulative within firms and technology-specific, generalizing existing evidence from the auto industry (Aghion et al., 2016). The third supports the key prediction that experienced incumbents are reluctant to innovate for new technologies given their expertise in old ones. I discuss these results in Section 5.5.

## 5.1 Data

The main dataset for the analysis is a panel of US public firms over 1980-2021, which includes measures of (i) each firm's patenting, both overall and within specific technologies; (ii) the aggregate stock of knowledge (patents) available to each firm, both overall and within specific technologies; and (iii) each firm's R&D expenditures. Below I summarize the main data sources, with additional details in Appendix D.

**Patents.** I use data on the set of all utility patents filed at the US Patent and Trademark Office (USPTO) after 1980 and granted through 2023. This dataset covers almost seven million patents and is made available through the USPTO's PatentsView platform. It includes each patent's title, abstract, assignees (initial owners), inventors, technology area, and citations made to other USPTO patents.

**Patent-Firm Matching.** Accurately grouping patents by firm is difficult, because assignee names are not standardized and are not adjusted to reflect changes in firm ownership through mergers and acquisitions.[28] I address these issues using the DISCERN 2.0 dataset, which employs an extensive matching process to identify all patents granted to US public firms from 1980 to 2021 (Arora et al., 2024). The resulting 1,865,633 patents are matched to 5,680 firms, and for tractability I restrict the analysis to the top 10% of firms by total number of patents. The final subset includes 1,659,998 patents matched to 568 firms. Restricting to public firms reduces the scope of the analysis, but it ensures an accurate matching of patents to firms and is common throughout the empirical literature on firm innovation (Hall et al., 2010).

**R&D Expenditures.** I use data on yearly nominal R&D expenditures from Compustat North America. I normalize by US GDP per capita each year to measure inflation-adjusted R&D expenditures in "scientist equivalent" units.

**New Technologies.** To group patents into technologies in a theory-consistent way, I start with the list of new technologies produced by Kalyani et al. (2023). They identify new technologies by extracting all two-word combinations that appear in the text of US patents over 1976-2014, excluding any combination found in a representative sample of pre-1970 American English text. Of the remaining "novel" combinations, the authors retain only the ones that appear in a sufficiently large number of citation-weighted patents. Finally, to ensure that these combinations refer to technologies instead of scientific concepts or problems to solve, the authors search for a Wikipedia page corresponding to each combination and verify that it describes a technology. Each of the resulting 1,148 two-word combinations is a *new technology*.

Kalyani et al. (2023) also define an "emergence year" for each new technology. This is meant to capture the time just before the new technology became prominent, but after it became available for firms to innovate. It is defined as the first year in which the number of citation-weighted patents that mention the technology (i) reaches a minimal threshold and (ii) grows by ten percent over each of the next five years. I use this emergence year in my analysis to mark the arrival of each new technology.

I identify a set of patents related to each new technology by searching for the two-word combination in the title and abstract of each patent. I include any patent that mentions the combination, as well as any patent that cites such a patent. Table 1 lists the top ten new technologies by total number of patents, along with their emergence years. The list primarily includes information and communication technologies ("data store," "code-division multiple

[28] PatentsView provides disambiguated assignee and inventor names meant to resolve the first issue, but my experience matching patents to pharmaceutical companies for the case study in Section 2 suggests that the PatentsView disambiguation protocol is still a work-in-progress.

Table 1: Top New Technologies by Total Patents

| Technology | Patents | Emergence Year |
|---|---|---|
| data store | 133,992 | 1990 |
| code-division multiple access | 107,638 | 1986 |
| memory address | 107,358 | 1992 |
| thermometer | 106,336 | 1991 |
| heat treating | 104,587 | 1983 |
| microsoft access | 100,097 | 1987 |
| error detection correction | 96,819 | 1986 |
| holographic optical element | 95,231 | 1992 |
| text-based user interface | 94,379 | 1994 |
| ion channel | 94,097 | 1992 |

*Notes:* Each technology's name corresponds to the title of its Wikipedia page, not the associated two-word combination.

access") along with biomedical technologies ("thermometer," "ion channel") and industrial technologies ("heat treating"), representative of the broader set of 1,148 technologies.

How well do these groups of patents capture the notion of a technology in the theory? The latter corresponds to (i) a shared set of production techniques and (ii) a shared base of knowledge used to improve those techniques. As Table 1 suggests, the new technologies identified in the patent data generally meet the first condition: Each new technology spans patents with similar applications, from data storage ("data store") to biological targets for drug treatment ("ion channel"). They also meet the second condition essentially by design: Any patent associated with a new technology must either mention it *or cite* one that does, so that it contributes directly to or builds on the technology's base of knowledge. To see that patents for a new technology are generally focused on similar topics, we can analyze how they are classified into technology areas by patent examiners. Each patent is assigned to potentially several of 132 "classes" and 672 "subclasses" defined in the Cooperative Patent Classification (CPC). These groupings are meant to identify all patents with similar technological content, with greater specificity for subclasses than classes. On average, I find that a share 0.58 of patents in a new technology belong to a *single class*, while a share 0.45 belong to a *single subclass*, so that patents associated with a new technology are highly concentrated within technology areas.[29]

[29] I do not use the CPC hierarchy to directly define technologies for two reasons. First, classes and subclasses

This method to group patents into technologies has several limitations. First, identifying an initial set of patents by keywords may include patents that simply mention the keywords but actually address other applications. I mitigate this concern by searching only in the title and abstract of each patent, which are more tightly focused on the patent's core applications. Second, Kalyani et al. (2023) explicitly identify *new* technologies that arrive over 1976-2014, but my theory focuses on the contrast between new and old technologies. In my empirical analysis, I proxy for old technologies with the set of all patents not associated with a new technology that were filed before its emergence year. Finally, some of the new technologies identified by Kalyani et al. (2023) have substantial overlap, such as "data store" and "microsoft access" (a database management system). This will attenuate my estimates of the technology specificity of knowledge, to the extent that I misclassify patents as unrelated to a given new technology. However, it may also artificially increase precision by inflating the number of new technologies above its true value.

## 5.2 Fact 1: Knowledge Accumulation within Firms

I first provide evidence consistent with knowledge accumulation within firms: A firm's current patenting is highly correlated with its previous patenting, holding fixed R&D expenditures, previous patenting by all firms, and the firm's latent propensity to patent. The regression is

$$Pat_{it} = \exp\left(\beta_1 \log\left(K_{it}^{\text{Firm}}\right) + \beta_2 \log\left(K_{it}^{\text{Agg}}\right) + \beta_3 \log(s_{it}) + \alpha' X_{it} + \varepsilon_{it}\right). \tag{32}$$

Here $Pat_{it}$ denotes the number of patents filed by firm $i$ in year $t$. I weight each patent by the number of forward citations it receives to better reflect the value of the underlying innovations.[30] The firm-level knowledge stock $K_{it}^{\text{Firm}}$ measures firm $i$'s accumulated internal knowledge at time $t$. I construct this stock from past patent flows by the perpetual inventory method, setting the depreciation rate $\nu$ to the standard value of 0.15:[31]

$$K_{it}^{\text{Firm}} = (1-\nu)K_{it-1}^{\text{Firm}} + Pat_{it-1}.$$

are still too broad to provide a compelling definition: Examples include class A61 "Medical or Veterinary Science; Hygiene" and subclass A61K "Preparations for Medical, Dental, or Toiletry Purposes", which include patents with a variety of applications. Second, many reasonably-defined technologies include patents assigned to many different classes and subclasses. For example, the two most common subclasses among mRNA therapy patents from Section 2 are A61K and C12N "Microorganisms or Enzymes; Compositions Thereof; Propogating, Preserving, or Maintaining Microorganisms; Mutation or Genetic Engineering; Culture Media."

[30] See Appendix D for additional details about the variable constructions.

[31] See Hall et al. (2005).

The aggregate knowledge stock $K_{it}^{\text{Agg}}$ measures the aggregate knowledge available to firm $i$ at time $t$. Since each firm draws on different knowledge depending on its area of focus, I construct $K_{it}^{\text{Agg}}$ in two steps. I first compute the knowledge stocks for each CPC subclass by the perpetual inventory method. I then compute $K_{it}^{\text{Agg}}$ as a weighted average of these stocks at each time $t$, with weights given by the distribution of firm $i$'s total patents across all CPC subclasses. Finally, I control for R&D expenditures $s_{it}$ and a vector of controls $X_{it}$ that includes separate fixed effects by year, firm, and the number of years for which the firm has been publicly listed.

The regression (32) is an empirical analogue to the quality evolution equation (8). I use the patent flow $Pat_{it}$ in place of the change in quality $\dot{q}_\theta(t)$, the firm-level knowledge stock $K_{it}^{\text{Firm}}$ in place of the quality $q_\theta(t)$, and $K_{it}^{\text{Agg}}$ in place of the aggregate knowledge stock $K_\theta(t)$. However, in this section the regression (32) is not technology-specific; it instead captures determinants of overall innovation. I adopt a log-linear functional form so that the coefficients of interest $\beta_1$, $\beta_2$, and $\beta_3$ may be interpreted as unit-free elasticities. The hypothesis that knowledge accumulates within firms implies $\beta_1 > 0$, so that a firm's past patenting has a larger impact on its current patenting than predicted by its effect on the aggregate knowledge stock $K_{it}^{\text{Agg}}$.

I estimate the regression (32) on the panel of public firms described in Section 5.1. To mitigate truncation issues in the patent data and in the unbalanced panel of firms, I restrict to the years 1985-2016, and I drop any observations for the first year when a firm becomes publicly listed.[32] I estimate directly by negative binomial regression and by linear regression (OLS) with dependent variable $\log(1+Pat_{it})$. The corresponding estimates are found in columns (1) and (2) of Table 2.

In each specification, I find that a firm's patenting is positively correlated with its past patenting ($K_{it}^{\text{Firm}}$) conditional on the other covariates. The estimates of the corresponding elasticity suggest that a 10% increase in a firm's internal knowledge stock $K_{it}^{\text{Firm}}$ is associated with a 7% increase in the firm's contemporaneous patenting $Pat_{it}$. This is consistent with the assumption of the theory that knowledge accumulates within firms ($\lambda > 0$), so that the productivity of a firm's R&D expenditures $s_{it}$ is increasing in its own past patenting $K_{it}^{\text{Firm}}$. The regression controls for many alternative explanations: For example, firm fixed effects ensure that this relationship is not driven by variation in firms' latent propensity to patent at all times, which would yield a positive correlation between past and current patenting. Controlling for R&D expenditures similarly ensures that the relationship is not driven by variation in the scale of each firm's R&D program. To the extent that R&D expenditures are elastic to *persistent* inno-

[32] Truncation issues arise in the patent data because I only observe *granted* patents, and there is often a multi-year lag between when a patent is filed and when it is granted. A similar problem arises for observed citations, which I address by normalizing forward citations both within year and across CPC classes (see Appendix D). A left truncation issue appears for firms because I only observe them after they become publicly listed, biasing my initial estimates of their internal knowledge stocks $K_{it}^{\text{Firm}}$ downward if they filed for patents before then.

Table 2: Regression Results: Firm Patenting

| | Overall Patenting | | Technology-Specific Patenting | |
|---|---|---|---|---|
| | (1) Neg. Binomial | (2) OLS | (3) Neg. Binomial | (4) OLS |
| $\log\left(K_{it}^{\text{Firm}}\right)$ | 0.696*** | 0.716*** | 0.286*** | 0.132*** |
| | (0.0329) | (0.0335) | (0.0617) | (0.0268) |
| $\log\left(K_{it}^{\text{Agg}}\right)$ | 0.297* | 0.248* | 0.554 | -0.0997 |
| | (0.125) | (0.119) | (0.554) | (0.147) |
| $\log\left(s_{it}\right)$ | 0.0778*** | 0.0951*** | 0.0961* | 0.0450** |
| | (0.0185) | (0.0203) | (0.0473) | (0.0162) |
| $\log\left(K_{i\theta t}^{\text{Firm}}\right)$ | | | 0.504*** | 0.397*** |
| | | | (0.0214) | (0.0285) |
| $\log\left(K_{\theta t}^{\text{Agg}}\right)$ | | | 0.337*** | 0.0919*** |
| | | | (0.0423) | (0.0122) |
| Observations | 12237 | 12228 | 79093 | 79092 |
| Firms | 555 | 546 | 257 | 256 |

*Significance: *** 0.01, ** 0.05, * 0.1*

*Notes:* The overall patenting regressions include fixed effects by year, firm, and the number of years for which the firm has been publicly listed. They also include dummy variables for zero internal knowledge stock $K_{it}^{\text{Firm}}$, zero aggregate knowledge stock $K_{it}^{\text{Agg}}$, and zero R&D expenditures $s_{it}$. The technology-specific patenting regressions additionally include fixed effects by technology and the time since emergence. They also add dummy variables for zero internal and aggregate technology-specific knowledge stocks $K_{i\theta t}^{\text{Firm}}$ and $K_{\theta t}^{\text{Agg}}$. All standard errors are clustered at the firm level.

vation opportunities, this control also rules out the explanation that firms simply patent more when they discover a valuable line of research.

Table 2 also shows that a firm's patenting is positively correlated with the measure of the aggregate knowledge $K_{it}^{\text{Agg}}$ available to the firm, which reflects positive spillovers across incumbents in the theory ($\sigma_I > 0$). The estimates of the corresponding elasticity imply that a 10% increase in the firm's external knowledge $K_{it}^{\text{Agg}}$ is associated with a 2.5-3% increase in contemporaneous patenting. I also find a positive elasticity of patenting to R&D expenditures of 8-9%, in line with previous estimates in the literature (e.g., Bloom et al., 2013).

## 5.3 Fact 2: Technology-Specific Knowledge

The evidence above supports the claim that knowledge accumulates within firms, but they say nothing about the extent to which this knowledge is *technology-specific*. To assess this second assumption of the theory, I consider a related regression for a firm's patenting in a given technology $\theta$:

$$Pat_{i\theta t} = \exp\Big(\beta_1 \log\left(K_{it}^{\text{Firm}}\right) + \beta_2 \log\left(K_{it}^{\text{Agg}}\right) + \beta_3 \log\left(s_{it}\right) \\ + \beta_4 \log\left(K_{i\theta t}^{\text{Firm}}\right) + \beta_5 \log\left(K_{\theta t}^{\text{Agg}}\right) + \alpha' X_{i\theta t} + \varepsilon_{i\theta t}\Big). \tag{33}$$

Here the dependent variable $Pat_{i\theta t}$ is the number of citation-weighted patents filed by firm $i$ for technology $\theta$ in year $t$. Regressors again include the firm's internal knowledge stock $K_{it}^{\text{Firm}}$, aggregate knowledge stock $K_{it}^{\text{Agg}}$, and R&D expenditures $s_{it}$, none of which are technology-specific. But I also add the technology-specific knowledge stocks $K_{i\theta t}^{\text{Firm}}$ and $K_{\theta t}^{\text{Agg}}$. Here $K_{i\theta t}^{\text{Firm}}$ is defined analogously to $K_{it}^{\text{Firm}}$ as the discounted sum of past citation-weighted patents by firm $i$ for technology $\theta$. The aggregate knowledge stock $K_{\theta t}^{\text{Agg}}$ is simply the discounted sum of all past patents for technology $\theta$ and does not vary across firms. The vector of controls $X_{i\theta t}$ includes the same fixed effects as in the first regression (32), in addition to fixed effects by technology $\theta$ and the number of years after the emergence of the technology. The hypothesis that knowledge is technology-specific implies $\beta_4 > \beta_1$ and $\beta_5 > \beta_2$, so that knowledge related to a given technology $\theta$ has a larger impact on a firm's current patenting for $\theta$ than generic knowledge.

I consider technologies $\theta$ with emergence years after 1990. I make a final restriction to ensure that the estimates of the regression (33) apply to firm-technology pairs such that the firm could plausibly innovate the technology: As described above for firms, for each technology $\theta$ I compute the distribution of its associated patents across CPC subclasses. I then exclude any firm-technology pairs for which the cosine similarity between the firm and technology patent distributions falls below 0.6. This excludes, for example, the pair of Microsoft and "microsatellite" (cosine similarity 0.44) while including the pair of Microsoft and "flash memory" (cosine similarity 0.61). Including more firm-technology pairs attenuates the coefficient estimates, as firms patent little for technologies outside their areas of focus.

The negative binomial and OLS estimates for the technology-specific regression (33) are found in columns (3) and (4) of Table 2. These estimates broadly demonstrate that knowledge is technology-specific: The estimated elasticities corresponding to the technology-specific knowledge stocks $K_{i\theta t}^{\text{Firm}}$ and $K_{\theta t}^{\text{Agg}}$ are significantly positive and generally larger than the elasticities corresponding to the generic knowledge stocks $K_{it}^{\text{Firm}}$ and $K_{it}^{\text{Agg}}$. For example, a 10% increase in a firm's past patenting for a technology $\theta$ correlates with a 4-5% increase in the firm's contemporaneous patenting for $\theta$, while the corresponding elasticity for the firm's generic past patenting is only 1.3-3%. Finally, note that the elasticity of patenting for technology $\theta$ with respect to R&D expenditures naturally attenuates relative to the firm-level specification (32):

$s_{it}$ is at best an imprecise proxy for the R&D expenditures specifically allocated by the firm to technology $\theta$, which I cannot observe.

## 5.4 Fact 3: Incumbents and New Technologies

When knowledge accumulates within firms and is technology-specific, the theory predicts that incumbent firms with extensive experience in existing technologies should be reluctant to innovate new ones, relative to less-experienced incumbents or entrants (Lemma 2). To test this prediction, I estimate several regressions that relate a firm's patenting for a new technology $\theta$ to the characteristics of the firm at the emergence time $T_\theta$:

$$\log(1+Pat_{i\theta}^{10}) = \beta_1 \log\left(K_{iT_\theta}^{\text{Firm}}\right) + \beta_2 \log\left(K_{iT_\theta}^{\text{Agg}}\right) + \beta_3 \log\left(s_{iT_\theta}\right) + \beta_4 \log\left(K_{i\theta T_\theta}^{\text{Firm}}\right) + \beta_5 \log\left(K_{\theta T_\theta}^{\text{Agg}}\right) + \alpha' X_{i\theta} + \varepsilon_{i\theta}. \tag{34}$$

Here $Pat_{i\theta}^{10}$ denotes the total number of citation-weighted patents filed by firm $i$ for technology $\theta$ in the ten years after the technology's emergence. The regressors include all firm-specific and aggregate knowledge stocks at the emergence time $T_\theta$, along with the firm's R&D expenditures $s_{iT_\theta}$. The vector of controls $X_{i\theta}$ includes fixed effects by firm $i$, emergence time $T_\theta$, and the number of years for which the firm has been publicly listed at the emergence time.[33] I also consider the alternative specification in which the outcome variable is the *share* of technology $\theta$ in all patents filed by firm $i$ in the ten years after emergence:

$$TechShare_{i\theta}^{10} \equiv \frac{Pat_{i\theta}^{10}}{\sum_{t=T_\theta}^{T_\theta+9} Pat_{it}}.$$

The regression (34) is similar to the previous one with technology-specific patenting (33), but with no panel dimension. It simply relates a firm's medium-run innovation for a new technology $Pat_{i\theta}^{10}$ to the firm's characteristics when that technology emerges. The coefficient of interest is $\beta_1$, which measures how the firm's initial knowledge stock at the time of emergence $K_{iT_\theta}^{\text{Firm}}$ predicts its subsequent patenting in the new technology $\theta$. When the dependent variable is $TechShare_{i\theta}^{10}$, the theory predicts that this coefficient should be negative: The empirical results documented thus far suggest that firm innovation decisions should be to some extent path-dependent. As a result, incumbents with greater internal knowledge stocks when a new technology $\theta$ emerges should devote a smaller share of their R&D resources to technology

[33] In this baseline specification, I exclude technology fixed effects so as to estimate the effect of the initial aggregate knowledge stock $K_{\theta T_\theta}^{\text{Agg}}$. The results in Table 3 are robust to the inclusion of technology fixed effects (see Appendix E).

Table 3: Regression Results: Technology Patenting after Emergence

| | Technology Patents | | Technology Patent Share | |
|---|---|---|---|---|
| | (1) Full Sample | (2) No Early Patents | (3) Full Sample | (4) No Early Patents |
| $\log(K^{\text{Firm}}_{iT_\theta})$ | -0.0828*** | -0.0796*** | -0.0024*** | -0.0014** |
| | (0.0315) | (0.0267) | (0.0009) | (0.0007) |
| Observations | 13,662 | 11,002 | 13,662 | 11,002 |

*Significance: *** 0.01, ** 0.05, * 0.1*

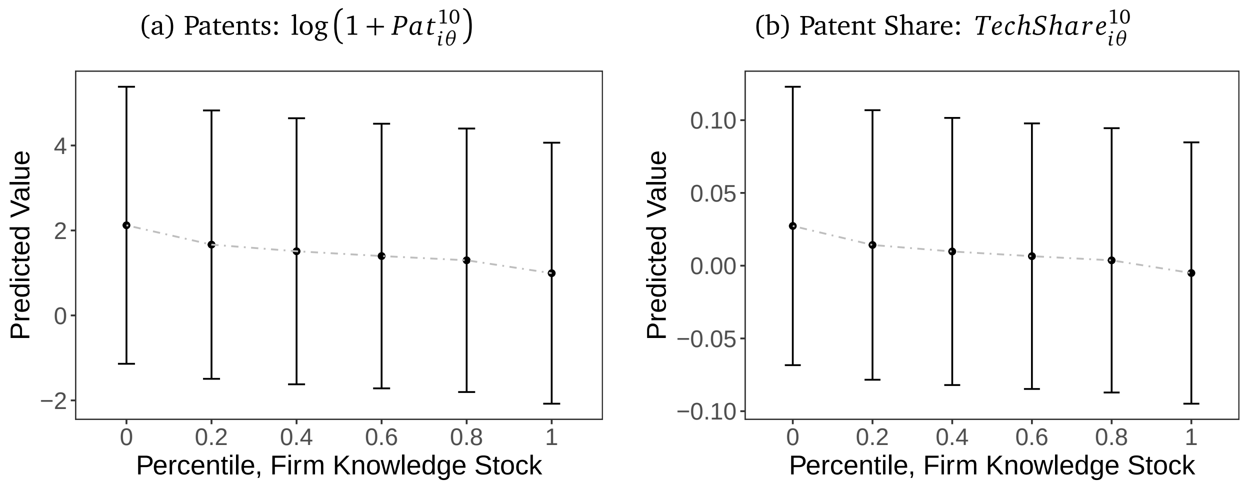
*Notes:* All regressions include fixed effects by firm, emergence year, and the number of years for which the firm has been publicly listed. They also include dummy variables for zero values of each of the knowledge stocks and R&D expenditures $s_{it}$. All standard errors are clustered at the firm level.

$\theta$, yielding in a smaller share for technology $\theta$ in the firm's total patents $TechShare^{10}_{i\theta}$. This need not imply that these firms patent *absolutely* less for technology $\theta$, because their greater research productivity may allow them to patent more in many different technologies.[34] Only when the path dependence force is particularly strong does the theory predict that $\beta_1$ should be negative when the dependent variable in the regression (34) is the transformed patent count $\log(1 + Pat^{10}_{i\theta})$.

I estimate the regression (34) by OLS for both dependent variables. For brevity, I display the estimates for $\beta_1$ in Table 3 and present the estimates for the remaining coefficients in Appendix E. Columns (1) and (3) provide the estimate $\hat{\beta}_1$ for the full sample of firms, while columns (2) and (4) exclude firms with any patenting for technology $\theta$ before the emergence year $T_\theta$. Consistent with firm-level path dependence, the estimates suggest a sizable negative effect of a firm's initial patenting experience on its subsequent patenting for new technologies: A 10% increase in a firm's initial internal knowledge stock $K^{\text{Firm}}_{iT_\theta}$ is associated with a 0.8% decrease in the firm's subsequent patents for the new technology. It also corresponds to a decrease of 0.1-0.2 percentage points in the technology's share of the firm's subsequent patents. To provide a sense of the effect size, Figure 3 plots predicted values from the regressions in columns (1) and (3) at different percentiles of the distribution of initial knowledge stocks $K^{\text{Firm}}_{iT_\theta}$, holding other covariates fixed at their mean values. Moving from the 10th percentile to the 90th percentile of the initial knowledge distribution reduces the firm's subsequent patents for a new technology by 53%; it also reduces the technology's share of the firm's subsequent patents by 1.6 percentage points, or approximately 27%. These effects suggest substantial path dependence in incumbent

[34]This force is absent in the theoretical model because all knowledge is technology-specific and firms cannot deploy more R&D resources as their research becomes more productive.

Figure 3: Predicted Values: Technology Patenting after Emergence




*Notes:* Figures 3(a) and 3(b) respectively display predicted values from the regressions in columns (1) and (3) of Table 3, with 95% confidence intervals. The firm-specific internal knowledge stock $\log\left(K^{\text{Firm}}_{it}\right)$ is evaluated at each quintile of its empirical distribution, with remaining covariates evaluated at their sample means. The distribution of $\log\left(K^{\text{Firm}}_{it}\right)$ ranges from -3.55 to 10.08, with median 4.59 and mean 4.26.

innovation, consistent with a fundamental prediction of the theory.

To assess how path dependence varies as a new technology matures, I estimate a dynamic version of the regression (34), allowing the effect of the initial internal knowledge stock $K^{\text{Firm}}_{iT_\theta}$ to vary over time:

$$\begin{aligned}\log(1 + Pat_{i\theta w}) = &\sum_{\tilde{w}=-2}^{5} \beta_{1\tilde{w}} \log\left(K^{\text{Firm}}_{iT_\theta}\right) \mathbb{1}_{w=\tilde{w}} + \beta_2 \log\left(K^{\text{Agg}}_{iT_\theta}\right) + \beta_3 \log\left(s_{iT_\theta}\right) \\ &+ \beta_4 \log\left(K^{\text{Firm}}_{i\theta T_\theta}\right) + \beta_5 \log\left(K^{\text{Agg}}_{\theta T_\theta}\right) + \alpha' X_{i\theta w} + \varepsilon_{i\theta w}.\end{aligned} \tag{35}$$

Here I group all years from 1980-2022 into three-year windows, and for each technology I denote the window containing the emergence time $T_\theta$ by $w = 0$. I let $Pat_{i\theta w}$ denote the firm's total citation-weighted patents for technology $\theta$ in window $w$. I group years into windows so that I can consider the alternative specification in which the outcome variable is the share of technology $\theta$ in all patents filed by firm $i$ in window $w$:

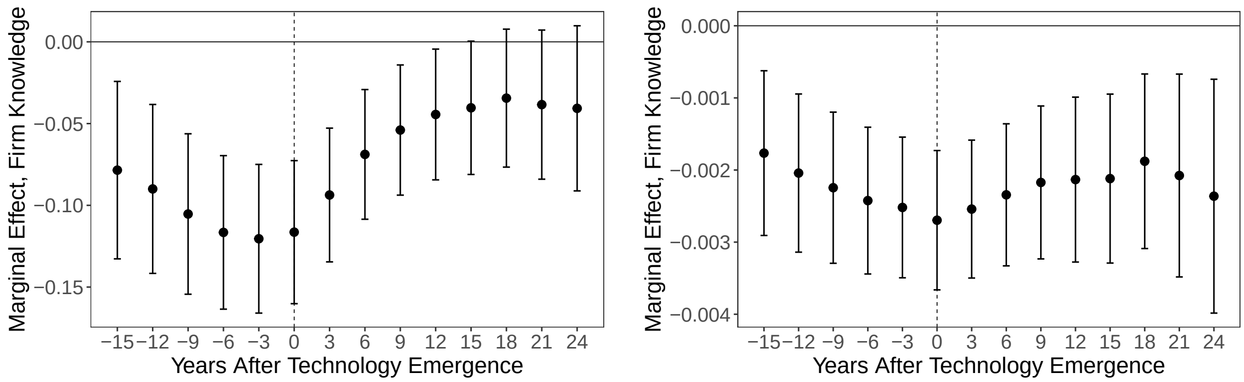
$$TechShare_{i\theta w} \equiv \frac{Pat_{i\theta w}}{Pat_{iw}}.$$

The timing of patent filings is noisy from year to year, so grouping years into windows reduces the number of observations dropped when the denominator of $TechShare_{i\theta w}$ is zero.

The regression (35) relates the firm's patenting after the emergence of a technology to its

Figure 4: Regression Results: Technology Patenting Over Time

(a) Patents: $\log(1 + Pat_{i\theta w})$ (b) Patent Share: $TechShare_{i\theta w}$




*Notes:* All regressions include fixed effects by firm-year, technology-window, and the number of years for which the firm has been publicly listed. They also include dummy variables for zero values of each of the knowledge stocks and R&D expenditures $s_{it}$. All standard errors are clustered at the firm level.

various knowledge stocks at the time of emergence, allowing for heterogeneous effects across time for the firm's generic internal knowledge stock $K^{\text{Firm}}_{iT_\theta}$. The vector of controls $X_{i\theta w}$ now includes fixed effects for each firm-year $it$, each technology-window $\theta w$, and the number of years for which the firm has been publicly listed at the time of emergence $T_\theta$. Note that I do *not* include firm-technology fixed effects, so the regression (35) does *not* correspond to an event study. Rather, the coefficients of interest $\beta_{1w}$ capture the *level differences* in patenting across firms with different internal knowledge stocks $K^{\text{Firm}}_{iT_\theta}$ at the emergence time. The key prediction of the theory is that these coefficients should be negative $\beta_{1w} < 0$ for windows $w$ near zero (i.e., near emergence).

I estimate the regression (35) by OLS for the dependent variables $\log(1 + Pat_{i\theta w})$ and $TechShare_{i\theta w}$. The corresponding estimates ("marginal effects") $\hat{\beta}_{1w}$ are plotted in Figure 4. These estimates provide clear evidence for firm-level path dependence: A firm with greater initial patenting experience patents strictly less for an emerging technology, both absolutely and relative to its total flow of patents.[35] Figure 4(a) shows that this negative effect on total patenting attenuates both before and after the emergence time. Experienced incumbents are particularly unlikely to innovate for an emerging technology, but after 10 years a firm's initial experience has no effect on its total patenting for the new technology. Figure 4(b) replicates the initial negative effect of the firm's internal knowledge stock for the share of the new technology in the firm's total patents. This negative effect displays the same "V shape" around emergence,

[35]As expected, the coefficient magnitudes are similar to those from the cross-sectional regressions in Table 3.

but it is more persistent. On average, an experienced incumbent always has a lower share of patents for a new technology than less-experienced firms.

### 5.5 Discussion

The empirical results presented in this section support the key assumptions and implications of my theory. The results in Section 5.2 provide evidence that firms accumulate knowledge through innovation that raises the productivity of their future R&D. Most microeconomic evidence on the return to R&D at the firm level ignores this dynamic channel (Hall et al., 2010; Bloom et al., 2013), which my results suggest is quantitatively important. Klette (1996) provides an early discussion of this issue, which I extend with an alternative regression framework for firm-level innovation and more comprehensive evidence about the knowledge accumulation channel.

The results in Section 5.3 additionally demonstrate that knowledge produced through innovation is highly technology-specific, indicating that firm innovation decisions are subject to the path-dependent and forward-looking forces highlighted in the theory (Lemma 2). These findings generalize existing work in environmental economics that show knowledge from dirty and clean innovation is technology-specific (Dechezleprêtre et al., 2014) and may generate firm-level path dependence (Aghion et al., 2016). I show that these properties are pervasive, extending to a variety of technologies in many industries. I also explicitly control for other firm-level determinants of innovation outcomes (like R&D expenditures) that could offer an alternative explanation for these findings.

Finally, the analysis in Section 5.4 tests the theory's core implication that firms with substantial expertise in existing technologies should be reluctant to innovate for new ones. The findings are consistent with the theory, and they provide novel evidence that experienced incumbents generally play a limited role in innovating emerging technologies. The pattern of innovation uncovered by this analysis is also consistent with the history of innovation in mRNA vaccines described in Section 2: Incumbents initially innovate less for a new technology, but their efforts intensify after younger firms develop the technology over several years. Lessons from the case study appear to generalize to the emergence of a variety of new technologies.

## 6 Quantitative Example

In this section, I provide a simple calibrated example of the model to assess how an increase in the concentration of scientists might affect equilibrium innovation and social welfare. I describe the calibration in Section 6.1, and I provide simulation results in Section 6.2.

## 6.1 Calibration

I calibrate all parameters of the model necessary to compute the equilibrium path of the knowledge stocks $K$ and the aggregate qualities $Q$. I first normalize the total mass of scientists to one, $S = 1$. I set the exit rate to $\delta = 0.075$, in line with recent estimates of the aggregate exit rate for US firms reported by Hopenhayn et al. (2022). I choose the discount rate $\rho = 0.075$ to ensure that the firms' total discount rate $\rho + \delta = 0.15$ falls in the middle of the range reported by Gormsen and Huber (2023).

To calibrate the parameters of the innovation process, I note that the estimates of the firm patenting equation (32) reported in Table 2 indicate that firms build substantially on their own past knowledge when innovating, with a smaller contribution from external knowledge. To keep the equilibrium dynamics simple, I consider the limiting case described in Section 4.3 and set the contribution of spillovers across incumbents to zero $\sigma_I = 0$. I then normalize $\lambda = 1$. I choose the parameters $\sigma_E = 0.4$ and $\eta_A = 0.026$ jointly to ensure that the initial BGP growth rate is $g_A^* = 0.02$ and that the tail parameter of the BGP firm-quality distribution $\xi_A^*$ matches the corresponding value for the US firm-productivity distribution as in Benhabib et al. (2021). Finally, I choose technology $B$'s productivity parameter $\eta_B = 0.03$ to ensure that this technology allows for 0.5% faster growth than technology $A$ along its BGP, $g_B^* = 0.025$.

## 6.2 Simulation Results

The calibrated model has a unique equilibrium, and it features a transition from technology $A$ to technology $B$ if and only if the aggregate knowledge ratio $\kappa(t)$ is initially above the threshold $\kappa^* \approx 0.872$. Figure 5 displays the trajectories of the aggregate knowledge ratio $\kappa(t)$ and the aggregate qualities $Q(t)$ for an equilibrium with a transition. The aggregate knowledge ratio $\kappa(t)$ is "U-shaped," initially decreasing before increasing asymptotically. But it remains above the threshold $\kappa^E$ at which entrants would instead innovate for technology $A$, ensuring that the economy converges to the BGP for technology $B$ in the long run. The aggregate quality $Q_B$ for technology $B$ increases with innovation, while the aggregate quality $Q_A$ for technology $A$ initially increases before declining to a positive limiting value due to firm reallocation.

To explore how the concentration of scientists $s$ can affect the equilibrium's propensity to transition, in Figure 6 I plot the equilibrium threshold $\kappa^*$ for different values of $s$ and $\rho$, keeping all other parameters of the model fixed. The figure indicates that an increase in the concentration of scientists increases the threshold $\kappa^*$ under the baseline calibration of the model, implying a *lower* propensity to develop the high-growth technology $B$ in equilibrium. The threshold $\kappa^*$ increases more rapidly with $s$ for higher values of the discount rate $\rho$, while it can instead decline with $s$ for lower values of $\rho$. These observations are exactly consistent with

Figure 5: Knowledge Ratio and Aggregate Quality Trajectories

(a) Knowledge Ratio $\kappa$ (b) Aggregate Qualities $Q_\theta$

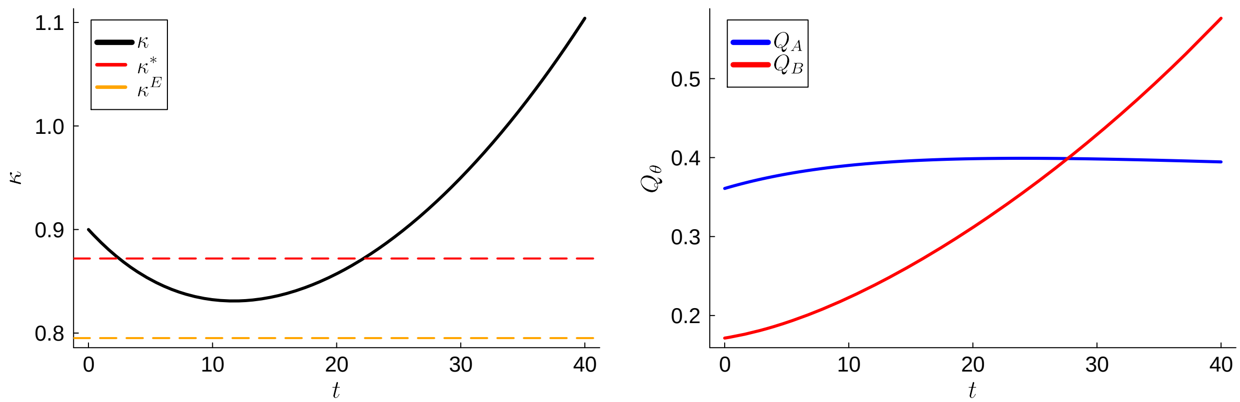



Figure 6: Equilibrium Threshold

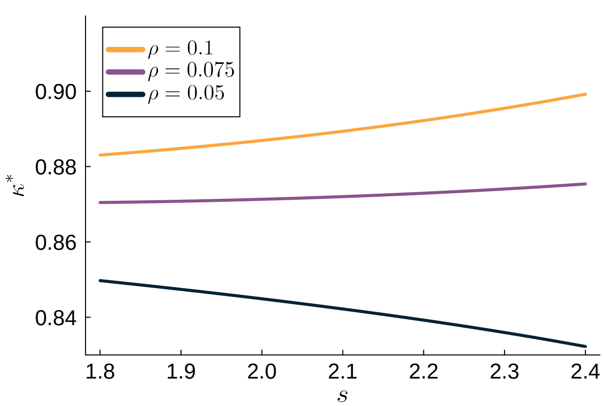
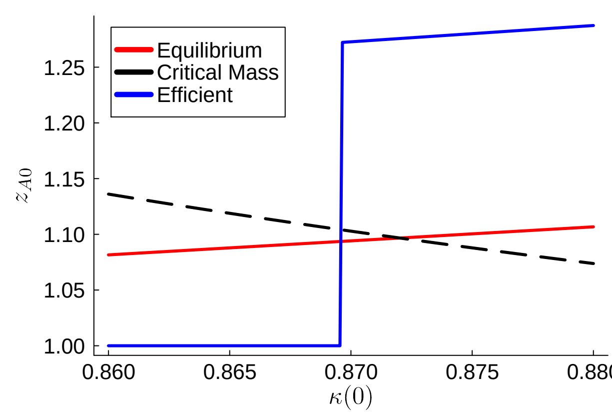



Figure 7: Equilibrium and Efficient Cutoffs $z_{A0}$



the trade-off between the composition and growth effects discussed in Section 4.3. When the discount rate $\rho$ is high, firms have short planning horizons, so any increase in future growth from greater concentration of scientists $s$ has a negligible impact on firm innovation decisions. However, it also produces a more skewed firm-quality distribution for technology $A$, leading a larger mass of initial incumbents to continue innovating for technology $A$.

Finally, it is straightforward to see that an increase in the concentration of scientists $s$ can actually reduce social welfare. In Figure 7, I plot the initial incumbent cutoff $z_{A0}$ chosen in equilibrium (red line) and by the social planner (blue line) as a function of the initial aggregate knowledge stock ratio $\kappa(0)$. This cutoff determines which initial incumbents innovate for technology $B$ ($z_A(0) \leq z_{A0}$), and it must be greater than the dashed "critical mass" line to ensure a transition to the BGP for technology $B$. Under the baseline calibration, the social planner always prefers to transition more often than in equilibrium. Consistent with Proposition 6 the efficient cutoff $\hat{z}_{A0}$ is always greater than the equilibrium cutoff $z^*_{A0}$ when a transition to technology $B$ is efficient.

An increase in $s$ raises the equilibrium transition threshold $\kappa^*$, represented in Figure 7 by the intersection between the red and black lines. If $\kappa(0)$ is found to the right of this intersection, the increase in $s$ can prevent a transition to technology $B$, strictly reducing the economy's long-run growth rate and social welfare. This holds despite the uniformly positive effects of an increase in $s$ on growth and welfare along the economy's BGPs.

# 7 Conclusion

In this paper, I presented a new model of innovation and firm dynamics to clarify a novel connection between market structure, the direction of innovation, and economic growth. The accumulation of knowledge within firms generates an endogenous comparative advantage for incumbents *in innovating* old technologies relative to entrants; the accumulation of public knowledge instead generates complementarities across firms. These features of the innovation process jointly imply that an industry's initial market structure affects its propensity to explore new technologies and hence its long-run growth. This mechanism has immediate relevance for the growing concentration of R&D in AI: Increasing concentration allows for faster progress for existing applications, but it also risks leaving valuable alternative innovation directions unexplored, as large incumbent firms develop AI according to their existing expertise. Empirical evidence from patenting for emerging technologies and a case study of mRNA vaccines supports the theory's mechanism and its key implication that incumbent firms are slow to innovate new technologies.

The broad implication of this paper is that focusing on complementarities between firms, not competition, provides a new perspective on the role different firms play in the innovation process. This basic idea touches on many aspects of innovation that would be interesting to explore in future research. For example, this paper considered only firms that are small relative to the size of the industry and labor market for scientists. But a "large" (granular) firm may internalize its effect on the innovation incentives for outside researchers, utilizing its size in the labor market to affect the direction of innovation in the industry. Similarly, complementarities in innovation may also explain (and suggest harm from) the investments large technology companies have made in small but innovative start-ups. More generally, these complementarities suggest a new theory of harm from industry consolidation even when competition is not threatened. Innovation or antitrust policies that promote diverse and vibrant innovation ecosystems may be essential for long-run growth.

# References


Acemoglu, D. (1998). Why do new technologies complement skills? directed technical change and wage inequality. *The quarterly journal of economics*, *113*(4), 1055–1089.

Acemoglu, D. (2002). Directed technical change. *The review of economic studies*, *69*(4), 781–809.

Acemoglu, D. (2011). *Diversity and technological progress* (tech. rep.). National Bureau of Economic Research.

Acemoglu, D. (2021). Ai's future doesn't have to be dystopian.' *Boston Review*, 23–29.

Acemoglu, D., & Cao, D. (2015). Innovation by entrants and incumbents. *Journal of Economic Theory*, *157*, 255–294.

Acemoglu, D., Aghion, P., & Zilibotti, F. (2006). Distance to frontier, selection, and economic growth. *Journal of the European Economic association*, *4*(1), 37–74.

Acemoglu, D., Aghion, P., Bursztyn, L., & Hemous, D. (2012). The environment and directed technical change. *American economic review*, *102*(1), 131–166.

Acemoglu, D., Akcigit, U., Hanley, D., & Kerr, W. (2016). Transition to clean technology. *Journal of political economy*, *124*(1), 52–104.

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

Aghion, P., Harris, C., Howitt, P., & Vickers, J. (2001). Competition, imitation and growth with step-by-step innovation. *The Review of Economic Studies*, *68*(3), 467–492.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The quarterly journal of economics*, *120*(2), 701–728.

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R., & Van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, *124*(1), 1–51.

Aghion, P., Bergeaud, A., De Ridder, M., & Van Reenen, J. (2024). Lost in transition: Financial barriers to green growth.

Ahmed, N., Wahed, M., & Thompson, N. C. (2023). The growing influence of industry in ai research. *Science*, *379*(6635), 884–886.

Akcigit, U., & Ates, S. T. (2021). Ten facts on declining business dynamism and lessons from endogenous growth theory. *American Economic Journal: Macroeconomics*, *13*(1), 257–298.

Akcigit, U., & Ates, S. T. (2023). What happened to us business dynamism? *Journal of Political Economy*, *131*(8), 2059–2124.

Akcigit, U., & Goldschlag, N. (2023). *Where have all the" creative talents" gone? employment dynamics of us inventors* (tech. rep.). National Bureau of Economic Research.

Akcigit, U., & Kerr, W. R. (2018). Growth through heterogeneous innovations. *Journal of Political Economy*, *126*(4), 1374–1443.

Alberer, M., Gnad-Vogt, U., Hong, H. S., Mehr, K. T., Backert, L., Finak, G., Gottardo, R., Bica, M. A., Garofano, A., Koch, S. D., et al. (2017). Safety and immunogenicity of a mrna rabies vaccine in healthy adults: An open-label, non-randomised, prospective, first-in-human phase 1 clinical trial. *The Lancet*, *390*(10101), 1511–1520.

Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *The American economic review*, *62*(5), 777–795.

Argente, D., Baslandze, S., Hanley, D., & Moreira, S. (2020). Patents to products: Product innovation and firm dynamics.

Arora, A., Belenzon, S., Cioaca, L., Sheer, L., Shin, H. M. (, & Shvadron, D. (2024). DISCERN 2: Duke Innovation & SCientific Enterprises Research Network. https://doi.org/10.5281/zenodo.13619821

Arrow, K. J. (1962). The economic implications of learning by doing. *The review of economic studies*, *29*(3), 155–173.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The economic journal*, *99*(394), 116–131.

Benhabib, J., Perla, J., & Tonetti, C. (2021). Reconciling models of diffusion and innovation: A theory of the productivity distribution and technology frontier. *Econometrica*, *89*(5), 2261–2301.

Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, *81*(4), 1347–1393.

Brock, W. A., & Durlauf, S. N. (1999). A formal model of theory choice in science. *Economic theory*, *14*, 113–130.

Brynjolfsson, E. (2023). The turing trap: The promise & peril of human-like artificial intelligence. In *Augmented education in the global age* (pp. 103–116). Routledge.

Christensen, C. M. (1997). *The innovator's dilemma: When new technologies cause great firms to fail*. Harvard Business Review Press.

Coase, R. (1937). The nature of the firm. *Economica*, *4*(16), 386–405.

Cunningham, C., Ederer, F., & Ma, S. (2021). Killer acquisitions. *Journal of political economy*, *129*(3), 649–702.

Dechezleprêtre, A., Märtin, R.-P., & Mohnen, M. (2014). *Knowledge spillovers from clean and dirty technologies*. Centre for Economic Performance.

de Soyres, F., Herrero, J. G.-C., Goernemann, N., Jeon, S., Lofstrom, G., & Moore, D. (2024). Why is the us gdp recovering faster than other advanced economies?

Dixit, A. K. (1993). *The art of smooth pasting* (Vol. 55). Taylor & Francis.

Dolgin, E. (2021). The tangled history of mrna vaccines. *Nature*, *597*(7876), 318–324.

Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research policy*, *11*(3), 147–162.

Dugoua, E., & Gerarden, T. (2023). *Induced innovation, inventors, and the energy transition* (tech. rep.). National Bureau of Economic Research.

Farrell, J., & Saloner, G. (1986). Installed base and compatibility: Innovation, product preannouncements, and predation. *The American economic review*, 940–955.

Gilbert, R. J., & Newbery, D. M. (1982). Preemptive patenting and the persistence of monopoly. *The American Economic Review*, 514–526.

Gormsen, N. J., & Huber, K. (2023). *Corporate discount rates* (tech. rep.). National Bureau of Economic Research.

Gote, V., Bolla, P. K., Kommineni, N., Butreddy, A., Nukala, P. K., Palakurthi, S. S., & Khan, W. (2023). A comprehensive review of mrna vaccines. *International journal of molecular sciences*, *24*(3), 2700.

Griliches, Z. (1998). *R&d and productivity: The econometric evidence*. University of Chicago Press.

Grossman, G. M., & Helpman, E. (1991). Quality ladders in the theory of growth. *The review of economic studies*, *58*(1), 43–61.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The nber patent citation data file: Lessons, insights and methodological tools.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, 16–38.

Hall, B. H., Mairesse, J., & Mohnen, P. (2010). Measuring the returns to r&d. In *Handbook of the economics of innovation* (pp. 1033–1082). Elsevier.

Hedestam, G. K., & Sandberg, R. (2023). Discoveries concerning nucleoside base modifications that enabled the development of effective mrna vaccines against covid-19.

Henderson, R. (1993). Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. *The RAND Journal of Economics*, 248–270.

Henderson, R., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9–30.

Hopenhayn, H., Neira, J., & Singhania, R. (2022). From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share. *Econometrica*, *90*(4), 1879–1914.

Howitt, P. (1999). Steady endogenous growth with population and r. & d. inputs growing. *Journal of Political Economy*, *107*(4), 715–730.

Igami, M. (2017). Estimating the innovator's dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998. *Journal of Political Economy*, *125*(3), 798–847.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, *108*(3), 577–598.

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

Jurowetzki, R., Hain, D., Mateos-Garcia, J., & Stathoulopoulos, K. (2021). The privatization of ai research (-ers): Causes and potential consequences–from university-industry interaction to public research brain-drain? *arXiv preprint arXiv:2102.01648*.

Kalyani, A., Bloom, N., Carvalho, M., Hassan, T., Lerner, J., & Tahoun, A. (2023). The diffusion of new technologies.

Karikó, K., Buckstein, M., Ni, H., & Weissman, D. (2005). Suppression of rna recognition by toll-like receptors: The impact of nucleoside modification and the evolutionary origin of rna. *Immunity*, *23*(2), 165–175.

Katz, M. L., & Shapiro, C. (1986). Technology adoption in the presence of network externalities. *Journal of political economy*, *94*(4), 822–841.

Kinch, M. (2018). *Between hope and fear*. Simon; Schuster.

Klette, T. J. (1996). R&d, scope economies, and plant performance. *The RAND Journal of Economics*, 502–522.

Klette, T. J., & Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of political economy*, *112*(5), 986–1018.

Klinger, J., Mateos-Garcia, J., & Stathoulopoulos, K. (2020). A narrowing of ai research? *arXiv preprint arXiv:2009.10385*.

König, M., Storesletten, K., Song, Z., & Zilibotti, F. (2022). From imitation to innovation: Where is all that chinese r&d going? *Econometrica*, *90*(4), 1615–1654.

Kuhn, T. (1970). The nature of scientific revolutions. *Chicago: University of Chicago*, *197*(0).

Lerner, J., & Seru, A. (2022). The use and misuse of patent data: Issues for finance and beyond. *The Review of Financial Studies*, *35*(6), 2667–2704.

Lucas, R. E., Jr, & Moll, B. (2014). Knowledge growth and the allocation of time. *Journal of Political Economy*, *122*(1), 1–51.

Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, *122*(3), 1103–1144.

Marshall, A. (1890). *Principles of economics*. Macmillan.

Milgrom, P., & Shannon, C. (1994). Monotone comparative statics. *Econometrica: Journal of the Econometric Society*, 157–180.

Murphy, S. L., Kochanek, K. D., Xu, J., & Arias, E. (2021). Mortality in the united states, 2021.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. harvard university press.

Reinganum, J. F. (1983). Uncertain innovation and the persistence of monopoly. *The American Economic Review*, *73*(4), 741–748.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of political economy*, *94*(5), 1002–1037.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, *98*(5, Part 2), S71–S102.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy*. harper.

Scott Morton, F., & Kyle, M. (2011). Markets for pharmaceutical products. In *Handbook of health economics* (pp. 763–823). Elsevier.

Yu, T. (2021). How scientists drew weissman (med'87, grs'87) and katalin karikó developed the revolutionary mrna technology inside covid vaccines. *Bostonia*. https://www.bu.edu/articles/2021/how-drew-weissman-and-katalin-kariko-developed-mrna-technology-inside-covid-vaccines/

# Appendix

## Contents

# A Proofs

This appendix provides proofs for results in the main text.

## A.1 Proofs for Section 3

**Proof of Lemma 1:** With linear, additively separable preferences over consumption sequences, the representative consumer's Euler equation and asset market clearing jointly imply $r(t) = \rho$. The intermediate price $p_\theta(q,t)$ must solve the maximization problem in the definition of flow profits (7), which implies a markup of $1/(1-\beta)$ over marginal cost: $p_\theta(q,t) = \gamma/(1-\beta)$. Given the final producer's intermediate demand curve, this implies the stated equations for intermediate quantities $x_\theta(q,t)$ and profits $\pi(q_\theta)$. Market-clearing for workers requires $L(t) = L$, and substituting the derived input quantities into the production function yields the state equation for output $Y(t)$. Market-clearing for goods then implies the corresponding equation for consumption $C(t)$. Finally, the wage for workers is recovered from the marginal product condition $w_L(t)L = \beta Y(t)$. ■

**Proof of Proposition 1:** The equations (17, 18) derived in the discussion before the Proposition characterize the growth rate $g_\theta^*$ and the relative quality distribution $H_\theta^*$. I reproduce these equations here:

$$g_\theta^* = \int (\lambda \sigma_E z_\theta + \sigma_I)\,\eta_\theta s\, dH_\theta^*(z_\theta), \tag{A1}$$

$$0 = -\left[(\lambda - g_\theta^*) z_\theta + \frac{\sigma_I}{\sigma_E}\right] h_\theta^*(z_\theta) + \delta N - H_\theta^*(z_\theta). \tag{A2}$$

The evolution equation (15) for $Q_\theta$ additionally implies

$$g_\theta^* = \frac{K_\theta(t)}{Q_\theta(t)}\left(g_\theta^* + \delta N \sigma_E\right) - \delta,$$

which can be rearranged to give the expression for $Q_\theta(t)/K_\theta(t)$ stated in the Proposition. The inequality $Q_\theta(t)/K_\theta(t) < 1$ follows immediately from this expression and the assumed upper bound on entrant spillovers (19), recalling that market-clearing for scientists requires $S = Ns$.

To solve the differential equation (A2) for $H_\theta^*$, define the values

$$a \equiv \lambda \eta_\theta s - g_\theta^*,$$
$$b \equiv \frac{\sigma_I}{\sigma_E}\eta_\theta s.$$

Then (A2) has the general solution

$$H_\theta^*(z_\theta) = \begin{cases} c\left(az_\theta + b\right)^{-\frac{\delta}{a}}\right) + N & a \neq 0, \\ c\exp\left(-\frac{\delta}{b}z_\theta\right) + N & a = 0. \end{cases}$$

The integration constant $c$ is determined by the condition that $H_\theta^*$ equals zero at the lower bound of its support. To determine the support, consider the quality of an intermediate introduced at $t = 0$. In a BGP with $K_\theta(t) = K_\theta(0)\exp\left(g_\theta^* t\right)$, we can integrate the evolution equation (8) for $q_\theta$ to find

$$q_\theta(t) = \begin{cases} \left[-\frac{b}{a} + \left(1 + \frac{b}{a}\right)\exp\left(at\right)\right]\sigma_E K_\theta(t) & a \neq 0, \\ (1 + bt)\sigma_E K_\theta(t) & a = 0. \end{cases}$$

Hence

$$z_\theta(t) = \begin{cases} -\frac{b}{a} + \left(1 + \frac{b}{a}\right)\exp\left(at\right) & a \neq 0, \\ 1 + bt & a = 0. \end{cases}$$

With $a > 0$ or $b > 0$, $z_\theta$ is strictly increasing over time and limits to infinity. Since firms exit at a finite rate, the BGP distribution $H_\theta^*$ must then have support $[1, \infty)$. In fact, I show below that $a = 0$ implies $b > 0$, so that $H_\theta^*$ has support $[1, \infty)$ when $a \geq 0$. If instead $a < 0$, I conjecture and verify below that $b > -a$. The expression above then implies that $z_\theta$ is strictly increasing over time and limits to the finite value $-b/a > 1$, so the distribution $H_\theta^*$ has bounded support $[1, -b/a]$.

In all cases, the integration constant $c$ is determined by the initial condition $H_\theta^*(1) = 0$. Solving this equation for $c$ and substituting yields

$$H_\theta^*(z_\theta) = \begin{cases} N\left[1 - \left(1 + \frac{a}{\delta}\frac{z_\theta - 1}{\frac{a+b}{\delta}}\right)^{-\frac{\delta}{a}}\right] & a \neq 0, \\ N\left[1 - \exp\left(-\frac{\delta}{b}(z_\theta - 1)\right)\right] & a = 0. \end{cases}$$

Hence $H_\theta^*$ is a generalized Pareto distribution with location parameter 1, shape parameter $\varphi_\theta^* = \frac{a+b}{\delta}$, and tail parameter $\xi_\theta^* = \frac{a}{\delta}$. Provided that $\xi_\theta^* < 1$, the mean of this distribution (scaled by $N$) is

$$\int z_\theta dH_\theta^*(z_\theta) = N\frac{\delta + b}{\delta - a} = N\frac{\delta + \frac{\sigma_I}{\sigma_E}\eta_\theta s}{\delta - (\lambda\eta_\theta s - g_\theta^*)}$$

Substituting into (A1), the growth rate $g_\theta^*$ must solve the fixed-point equation

$$g_\theta^* = \left[\lambda \frac{\sigma_E \delta + \sigma_I \eta_\theta s}{\delta - (\lambda \eta_\theta s - g_\theta^*)} + \sigma_I\right] \eta_\theta S.$$

Multiplying through by the denominator on the right side yields a quadratic equation in $g_\theta^*$, and the solution is given by the expression for $g_\theta^*$ in the Proposition.[36] With this expression, we immediately observe that the assumption of positive spillovers from innovation $\lambda\sigma_E + \sigma_I > 0$ ensures that $g_\theta^*$ is positive, and in fact large enough to guarantee $\xi_\theta^* < 1$: This inequality holds if and only if $g_\theta^* > \lambda \eta_\theta s - \delta$. This is immediate if $\lambda \eta_\theta s < \delta$, and otherwise it follows from $\lambda\sigma_E + \sigma_I > 0$ after substituting the expression for $g_\theta^*$ from the Proposition.

It remains to verify that $a = 0$ implies $b > 0$ and $b > -a$. For the former, note that $b \geq 0$, with equality if and only if $\sigma_I = 0$. If both $a = 0$ and $\sigma_I = 0$, the fixed-point equation for $g_\theta^*$ above simplifies to $s = \sigma_E S$, contradicting the assumed upper bound on entrant spillovers (19). To see that $b > -a$, note that this holds if and only if $\lambda + \frac{\sigma_I}{\sigma_E}\eta_\theta s > g_\theta^*$. After substituting the expression for $g_\theta^*$ from the Proposition, direct calculation shows that this inequality is also implied by the assumed upper bound on entrant spillovers (19). Note that this implies that the shape parameter $\varphi_\theta^*$ is positive.

To conclude, note that in all cases the assumed lower bound for $\rho$ (20) ensures that the consumer's transversality condition (3) holds. Moreover, the expression for $q_\theta(t)$ derived avove implies an asymptotic growth rate of $\lambda \eta_\theta s$ for flow profits, so the assumed lower bound (14) on the exit rate $\delta$ ensures that the value function $V(q,t)$ remains finite. ■

**Proof of Corollary 1:** I first state the complete set of comparative statics for the tail parameter $\xi_\theta^*$: The tail parameter $\xi_\theta^*$ is strictly decreasing in $\sigma_E$, $\sigma_I$, and $S$, and it is strictly increasing in $\lambda$ and $s$. If $\sigma_I$ is sufficiently close to zero, $\xi_\theta^*$ is strictly increasing in $\eta_\theta$ and strictly decreasing in $\delta$, while the opposite holds for $\sigma_I$ sufficiently large.

The comparative statics with respect to $\sigma_E$, $\sigma_I$, and $S$ follow immediately by the formula for $\xi_\theta^*$ in Proposition 1. For the comparative static with respect to $s$, we can differentiate the equation for $g_\theta$ to find

$$\frac{\partial g_\theta^*}{\partial s} = \frac{\lambda \eta_\theta}{2}\left[1 - \frac{\frac{\delta - \lambda\eta_\theta s - \sigma_I \eta_\theta S}{2}}{\sqrt{\left(\frac{\delta - \lambda\eta_\theta s - \sigma_I \eta_\theta S}{2}\right)^2 + (\lambda\sigma_E + \sigma_I)\delta \eta_\theta S}}\right].$$

With $\lambda\sigma_E + \sigma_I > 0$ by assumption, the term in brackets is bounded strictly below 2. Thus

[36] The lower root of the quadratic equation is negative and hence extraneous.

$\partial g_\theta^*/\partial s < \lambda\eta_\theta$, and $\xi_\theta^*$ is strictly increasing in $s$. A similar calculation implies $\partial g_\theta^*/\partial\lambda < \eta_\theta s$, making use of the assumed upper bound to spillovers on entrants (19). This implies that $\xi_\theta^*$ is strictly increasing in $\lambda$. Finally, for $\eta_\theta$,

$$\frac{\partial g_\theta^*}{\partial \eta_\theta} = \frac{\lambda s + \sigma_I S}{2} + \frac{-\frac{\lambda s+\sigma_I S}{2}\frac{\delta-\lambda\eta_\theta s-\sigma_I\eta_\theta S}{2} + (\lambda\sigma_E+\sigma_I)\delta S}{\sqrt{\left(\frac{\delta-\lambda\eta_\theta s-\sigma_I\eta_\theta S}{2}\right)^2 + (\lambda\sigma_E+\sigma_I)\delta\eta_\theta S}}.$$

If $\sigma_I = 0$, this value is below $\lambda s$ if and only if $s > \sigma_E S$, which holds by assumption (19). However, this value also limits to infinity as $\sigma_I \to \infty$. Thus $\xi_\theta^*$ is increasing in $\eta_\theta$ for $\sigma_I$ small, while it is decreasing in $\eta_\theta$ for $\sigma_I$ large. Since $g_\theta^*$ is linearly homogeneous in $(\delta, \eta_\theta)$, the opposite comparative statics hold for $\delta$: $\xi_\theta^*$ is decreasing in $\delta$ for $\sigma_I$ small, while it is increasing in $\delta$ for $\sigma_I$ large.

Finally, note from the expression for the shape parameter $\varphi_\theta^*$ given in Proposition 6 that

$$\varphi_\theta^* = \xi_\theta^* + \frac{\sigma_I}{\sigma_E}\frac{\eta_\theta s}{\delta}.$$

The results above them imply that $\varphi_\theta^*$ is strictly increasing in $s$. ■

## A.2 Proofs for Section 4

**Proof of Proposition 2:** We can directly integrate the quality evolution equation (8) to find that for $t' \geq t$,

$$q_\theta(t') = \exp\left(\lambda\eta_\theta \int_t^{t'} s_\theta(q(\tau),\tau)d\tau\right) q_\theta$$
$$+ \sigma_I\eta_\theta \int_t^{t'} s_\theta(q(t''),t'')K_\theta(t'')\exp\left(\lambda\eta_\theta \int_{t''}^{t'} s_\theta(q(\tau),\tau)d\tau\right)dt''.$$

Holding fixed the allocation of scientists $s(q(\tau),\tau)$ at $\tau \geq t$, we have

$$\frac{\partial q_\theta(t')}{\partial q_\theta} = \exp\left(\lambda\eta_\theta \int_t^{t'} s_\theta(q(\tau),\tau)d\tau\right).$$

Given an optimal allocation of scientists $s(q(\tau),\tau)$ at $\tau \geq t$ incumbent's value at $t$ is

$$V(q,t) = \bar{\pi}\int_t^\infty \exp\big(-(\rho+\delta)(t'-t)\big)[q_A(t')+q_B(t')]dt'.$$

The Envelope Theorem then implies

$$\frac{\partial V(q,t)}{\partial q_\theta} = \bar{\pi}\int_t^\infty \exp\big(-(\rho+\delta)(t'-t)\big)\frac{\partial q_\theta(t')}{\partial q_\theta}dt' = \bar{\pi}\Psi_\theta(t).$$

The remaining results follow immediately. ■

**Proof of Proposition 3:** See Appendix B.2. ■

**Proof of Proposition 4:** See Appendix B.3. ■

**Proof of Proposition 5:** See the proof of Proposition C.4 in Appendix C. ■

**Proof of Proposition 6:** See the proof of Proposition C.5 in Appendix C. ■

# B Monotone Equilibria

This appendix characterizes and proves the existence of monotone equilibria.

## B.1 Optimal Stopping: Proof of Lemma 2

To facilitate the analysis of monotone equilibria, I first prove several properties of solutions to the firm's problem (11) given an arbitrary but differentiable knowledge stock trajectory $[K(t)]_t$. Define the notation

$$\begin{aligned} k_\theta(t) &\equiv \lambda q_\theta(t) + \sigma_I K_\theta(t), \\ \Psi_\theta(T) &\equiv \int_T^\infty \exp\left(-\int_T^t [\rho + \delta - \lambda \eta_\theta s_\theta(q(\tau), \tau)]\, d\tau\right) dt, \\ \dot{k}_\theta(T-) &\equiv \lim_{t \uparrow T} \dot{k}(t). \end{aligned}$$

Note that $\Psi_\theta(T)$ satisfies the bounds

$$\Psi_\theta(T) \in \left[\frac{1}{\rho+\delta}, \frac{1}{\rho+\delta-\lambda\eta_\theta s}\right].$$

The following lemma provides a stopping time representation of solutions to the firm's problem (11), implying Lemma 2 in Section 4.1.

**Lemma B.1.** *Suppose $[K(t)]_t$ is differentiable, and suppose a piecewise-continuous solution $[s_\theta(q(t), t)]_t$ to (11) given initial qualities $q(t_0)$ at time $t_0 \geq 0$. Then without loss of generality $s_\theta(q(t), t) \in \{0, s\}$ for all $t \geq t_0$, and there exists a sequence of stopping times $t_0 < T_1 \leq T_2 \leq \ldots$ such that*

*(i) if $s_\theta(q(t), t) = s$ for $t \in [T_m, T_{m+1})$, then $s_\theta(q(t), t) = 0$ for $t \in [T_{m+1}, T_{m+2})$;*

*(ii) every positive stopping time $T_m > 0$ satisfies the smooth-pasting condition*

$$k_B(T_m)\eta_B \Psi_B(T_m) = k_A(T_m)\eta_A \Psi_A(T_m);$$

*(iii) every positive stopping time $T_m > 0$ at which the firm transitions from A to B satisfies the necessary second-order condition*

$$\frac{\dot{k}_B(T_m-)}{k_B(T_m)} + \rho + \delta - \lambda\eta_B s - \frac{1}{\Psi_B(T_m)} \geq \frac{\dot{k}_A(T_m-)}{k_A(T_m)} + \rho + \delta - \frac{1}{\Psi_A(T_m)}; \text{ and}$$

*(iv) every positive stopping time $T_m > 0$ at which the firm transitions from B to A satisfies the*

*necessary second-order condition*

$$\frac{\dot{k}_B(T_m-)}{k_B(T_m)}+\rho+\delta-\frac{1}{\Psi_B(T_m)}\leq\frac{\dot{k}_A(T_m-)}{k_A(T_m)}+\rho+\delta-\lambda\eta_A s-\frac{1}{\Psi_A(T_m)}.$$

**Proof of Lemma B.1:** The optimality of corner allocations $s_\theta(q(t),t)\in\{0,s\}$ follows immediately from the linearity of the evolution equation (8) in $s_\theta(q,t)$. Hence any solution can be identified with a sequence of (potentially infinite) stopping times $t_0<T_1\leq T_2\leq\ldots$ that prescribe when the firm should reverse its innovation direction.

Consider any positive stopping time $T_m>0$, and suppose without loss of generality that the firm transitions from $A$ to $B$ at $T_m$. We can directly integrate the quality evolution equation (8) to find that for $t\geq T_m$,

$$\begin{aligned}q_\theta(t)=&\exp\left(\lambda\eta_\theta\int_{T_m}^{t}s_\theta(q(\tau),\tau)d\tau\right)q_\theta(T_m)\\&+\sigma_I\eta_\theta\int_{T_m}^{t}s_\theta(q(t'),t')K_\theta(t')\exp\left(\lambda\eta_\theta\int_{t'}^{t}s_\theta(q(\tau),\tau)d\tau\right)dt'.\end{aligned}$$

The incumbent's value at $t_0$ is

$$V(q(t_0),t_0)=\bar{\pi}\int_{t_0}^{\infty}\exp\left(-(\rho+\delta)t\right)\left[q_A(t)+q_B(t)\right]dt.$$

The stopping time $T_m>0$ must satisfy the interior first-order condition

$$0=\frac{\partial}{\partial T_m}\frac{V(q(t_0),t_0)}{\bar{\pi}}=\int_{T_m}^{\infty}\exp\left(-(\rho+\delta)t\right)\left[\frac{\partial q_A(t)}{\partial T_m}+\frac{\partial q_B(t)}{\partial T_m}\right]dt.$$

For $t\geq T_m$, we can directly calculate

$$\frac{\partial q_\theta(t)}{\partial T_m}=\exp\left(\lambda\eta_\theta\int_{T_m}^{t}s_\theta(q(\tau),\tau)d\tau\right)\left[\dot{q}_\theta(T_m-)-\dot{q}_\theta(T_m+)\right].$$

Here $\dot{q}_\theta(T_m-)$ denotes the evolution of $q_\theta$ just before $T_m$, while $\dot{q}_\theta(T_m+)$ denotes the evolution just after $T_m$. The interior first-order condition for $T_m$ simplifies to

$$0=\sum_{\theta\in\{A,B\}}\left[\dot{q}_\theta(T_m-)-\dot{q}_\theta(T_m+)\right]\Psi_\theta(T_m).$$

Now given that the firm transitions from $A$ to $B$ at $T_m$< we have

$$\dot{q}_A(T_m-) - \dot{q}_A(T_m+) = k_A(T_m)\eta_A s,$$
$$\dot{q}_\theta(T_m-) - \dot{q}_\theta(T_m+) = -k_B(T_m)\eta_B s.$$

Hence the interior first-order condition is exactly the smooth-pasting condition stated in the Lemma:

$$k_B(T_m)\eta_B\Psi_B(T_m) = k_A(T_m)\eta_A\Psi_A(T_m).$$

The second-order necessary condition for $T_m$ requires that the left side of this equation be weakly increasing relative to the right side just before $T_m$. Log differentiating yields the necessary condition stated in the Lemma. The corresponding necessary condition for a transition from $B$ to $A$ is derived analogously. ■

## B.2 Equilibrium Characterization: Proof of Proposition 3

In this section, I prove Proposition 3 and characterize all monotone equilibria converging to the BGP for technology $B$. I then state the analogous result for monotone equilibria converging to the BGP for $A$. The proof of this result is almost identical to that for Proposition 3, so I only sketch the differences.

### B.2.1 Equilibrium Converging to $B$: Proof of Proposition 3

**Step 0: $T_E$ existence.** As noted before the statement of Proposition 3, for any monotone equilibrium converging to $B$ there exists a first time $T_E \geq 0$ at which entrants begin innovating immediately for technology $B$. The existence of this time is immediate: If the equilibrium converges to $B$, all entrants must eventually innovate for technology $B$ exclusively. Hence there exists a smallest time $T_E < \infty$ after which this holds, and monotonicity implies that all entrants at $t \in [0, T_E)$ initially innovate for technology $A$. Any equilibrium with $T_E > 0$ is essentially an equilibrium with $T_E = 0$, but shifted forward in time. The distribution of relative qualities for $A$ at $T_E$ may stochastically dominate the initial distribution $H_{A0}$ if the latter is not the BGP distribution $H_A^*$. But this simply implies incumbents are less willing to innovate for technology $B$ at $T_E$, so whenever an equilibrium transitioning to $B$ with $T_E > 0$ exists an equilibrium transitioning to $B$ with $T_E = 0$ must also exist. This justifies my focus on the latter in the main text and throughout the proof below.

**Step 1: $\chi(t)$ existence.** To prove the existence of the cutoff $\chi(t)$, fix a time $t_0 \geq 0$ and a firm with qualities $q_\theta(t_0)$. Given trajectories for the knowledge stocks $[K(t)]_t$ and the firm's allocation of scientists $[s_\theta(q(t),t)]_{\theta,t}$, we can directly integrate the quality evolution equation (8) to find

$$\begin{aligned} q_\theta(t) = &\exp\left(\lambda\eta_\theta\int_{t_0}^{t} s_\theta(q(\tau),\tau)d\tau\right)q_\theta(t_0) \\ &+ \sigma_I\eta_\theta\int_{t_0}^{t} s_\theta(q(t'),t')K_\theta(t')\exp\left(\lambda\eta_\theta\int_{t'}^{t} s_\theta(q(\tau),\tau)d\tau\right)dt'. \end{aligned}$$

This value is convex and supermodular in $[s_\theta(q(\tau),\tau)]_{\tau\in(t_0,t]}$, with increasing differences in $q_\theta(t_0)$ and $[s_\theta(q(\tau),\tau)]_{\tau\in(t_0,t]}$. Since flow profits are linear in qualities, the objective of the firm's problem (11) must also be supermodular in $[s_A(q(t),t)]_t$, with increasing differences in $q_A(t_0)$ and $[s_A(q(t),t)]_t$. Theorem 4 of Milgrom and Shannon (1994) then implies that the optimal value of $s_A(q(t),t)$ is non-decreasing in $q_A(t_0)$ at each $t \geq t_0$.

This observation immediately implies the existence of the cutoff $\chi(t)$: First setting $t_0 = 0$, we can observe that initial incumbents with higher relative qualities $z_A(0)$ transition later to technology $B$ (if at all). The expression for $q_\theta(t)$ above implies that $q_A(t)$ is strictly increasing in $z_A(0)$ for these firms, so that the ratio

$$\frac{k_B(t)}{k_A(t)} = \frac{\lambda q_B(t) + \sigma_I K_B(t)}{\lambda q_A(t) + \sigma_I K_A(t)}$$

is always strictly decreasing in $z_A(0)$. We can simply define $\chi(t)$ as the least upper bound of this ratio across all initial incumbents still innovating for technology $A$.

**Step 2: $\chi(t)$ characterization.** The construction above implies that $\chi(0)$ must be weakly above $\frac{K_B(0)}{K_A(0)}$. To describe the evolution of $\chi(t)$, note that at $t \geq 0$ new entrants permanently innovate for technology $B$, while all remaining firms must choose whether to continue innovating for $A$ or transition back to $B$. Lemma B.1 implies that any firm transitioning to $B$ at $T$ must satisfy the smooth-pasting condition

$$\frac{k_B(T)}{k_A(T)} = \frac{\eta_A}{\eta_B}\frac{\rho+\delta-\lambda\eta_B s}{\rho+\delta}, \tag{B1}$$

with the left side increasing at $T$. I claim that these conditions are met at most once for each firm that initially innovates for technology $A$.

To prove this fact, it is helpful to first derive evolution equations for the knowledge stocks $K(t)$. Let $T(z_A) \geq 0$ denote the transition time for a firm with initial relative quality $z_A(0)$. Let

$\tilde{z}_A(t)$ denote the inverse of this function, taking value 1 outside its image. Given the initial distribution of relative qualities $H_{A0}$, the evolution equation (9) for $K_A$ can be written

$$\dot{K}_A(t) = \exp(-\delta t) \int_{\tilde{z}_A(t)}^{\infty} [\lambda q_A(t; z_A) + \sigma_I K_A(t)] \eta_A s dH_{A0}(z_A),$$

where $q_A(t; z_A)$ denotes the quality for a firm with initial relative quality $z_A(0)$. Let $N_\theta(t)$ denote the total mass of firms innovating for technology $\theta$ at $t$, where

$$N_A(t) = \exp(-\delta t)[N - H_{A0}(\tilde{z}_A(t))] \quad \text{and} \quad N_B(t) = N - N_A(t).$$

Then we can differentiate again to derive the second-order differential equation

$$\begin{aligned} \ddot{K}_A(t) = &-[\delta - (\lambda + \sigma_I N_A(t)) \eta_A s] \dot{K}_A(t) \\ &- \frac{\rho + \delta}{\rho + \delta - \lambda \eta_B s} [\lambda \sigma_E K_B(0) + \sigma_I K_B(t)] \eta_B s \exp(-\delta t) h_{A0}(\tilde{z}_A(t)) \dot{\tilde{z}}_A(t). \end{aligned} \tag{B2}$$

Turning to technology $B$, we can similarly write the evolution equation (9) as

$$\begin{aligned} \dot{K}_B(t) = \delta N &\int_{T_E}^{t} \exp(-\delta(t - t_0)) [\lambda q_B(t; t_0) + \sigma_I K_B(t)] \eta_B s dt_0 \\ &+ \int_{1}^{\tilde{z}_A(t)} \exp(-\delta t) [\lambda q_B(t; z_A) + \sigma_I K_B(t)] \eta_B s dH_{A0}(z_A), \end{aligned}$$

where I abuse notation by letting $q_B(t; t_0)$ denote the quality for a firm that entered at time $t_0 \geq 0$ and $q_B(t; z_A)$ the quality of a firm with initial relative quality $z_A(0)$. Differentiating again yields

$$\begin{aligned} \ddot{K}_B(t) = &-[\delta - (\lambda + \sigma_I N_B(t)) \eta_B s] \dot{K}_B(t) + \delta N (\lambda \sigma_E + \sigma_I) \eta_B s K_B(t) \\ &+ [\lambda \sigma_E K_B(0) + \sigma_I K_B(t)] \eta_B s \exp(-\delta t) h_{A0}(\tilde{z}_A(t)) \dot{\tilde{z}}_A(t). \end{aligned} \tag{B3}$$

The evolution equations (B2, B3) have two important implications for the remainder of the proof. First, (B2) easily implies that $\ddot{K}_A(t) < 0$ because of the assumed lower bound (14) on $\delta$. Second, (B3) implies that the growth rate $g_B(t) \equiv \frac{\dot{K}_B(t)}{K_B(t)}$ is strictly increasing to its BGP value $g_B^*$. This growth rate evolves according to

$$\begin{aligned} \dot{g}_B(t) &= \frac{\ddot{K}_B(t)}{K_B(t)} - g_B(t)^2 \\ &\geq \delta N(\lambda \sigma_E + \sigma_I) \eta_B s - [\delta - (\lambda + \sigma_I N_B(t)) \eta_B s + g_B(t)] g_B(t). \end{aligned}$$

The right side of this inequality must be strictly positive for $t < \infty$: Define a new function $\check{g}(t)$ such that $\check{g}(0) = g_B(0)$, but with evolution equation

$$\dot{\check{g}}(t) = \delta N(\lambda\sigma_E + \sigma_I)\eta_B s - [\delta - (\lambda + \sigma_I N_B(0))\eta_B s + \check{g}_B(t)]\,\check{g}(t).$$

Direct calculation implies $\dot{\check{g}}(0) > 0$, making use of the assumed upper bound (19) on spillovers to entrants $\sigma_E$. With $\dot{N}_B(t) > 0$, we then have $\dot{\check{g}}(t) > 0$ and hence $\dot{g}_B(t) > 0$.

Returning to the smooth-pasting condition (B1), consider any firm innovating for technology $A$ at $t = 0$. If $\lambda\eta_A s > g_B^*$, it is easy to see that the firm never innovates for technology $B$: Lemma B.1 implies that the necessary second-order condition corresponding to the smooth-pasting condition (B1) is

$$\frac{\sigma_I K_B(T)}{k_B(T)} g_B(T) \geq \lambda\eta_A s + \frac{\sigma_I K_A(T)}{k_A(T)} g_A(T). \tag{B4}$$

But this inequality must be violated when $\lambda\eta_A s > g_B^*$, because the growth rate $g_B(T)$ is strictly lower than its BGP value $g_B^*$ and $\sigma_I K_B(T) < k_B(T)$ by definition.[37]

Suppose instead $\lambda\eta_A s < g_B^*$ and $\sigma_I > 0$. I claim that any firm innovating for technology $A$ at $t = 0$ must have a finite stopping time $T > 0$ after which it permanently innovates for technology $B$. The conditions $\lambda\eta_A s < g_B^*$ and $\sigma_I > 0$ easily imply that such a stopping time exists: With $g_B(t) \uparrow g_B^*$ and $g_A(t) \downarrow 0$, the technology $B$ research productivity $k_B(t)$ eventually grows faster than the technology $A$ research productivity $k_A(t)$. Moreover, it is straightforward to see that $T$ is the unique solution to the smooth-pasting condition (B1) that satisfies the corresponding second-order condition (B4). While the firm innovates for technology $A$, the left side of this condition is strictly increasing because both $g_B(t)$ and $\sigma_I K_B(T)/k_B(T)$ are strictly increasing. The right side satisfies

$$\frac{\partial}{\partial T}\frac{\dot{K}_A(T)}{k_A(T)} = \frac{\dot{K}_A(T)}{k_A(T)}\left(\frac{\ddot{K}_A(T)}{\dot{K}_A(T)} - \frac{\lambda\eta_A s + \sigma_I \dot{K}_A(T)}{k_A(T)}\right)$$

This value is negative because $\ddot{K}_A(T) < 0$. Thus for any firm that innovates for technology $A$ at $t = 0$, the smooth-pasting and necesssary second-order conditions for a permanent transition to technology $B$ are satisfied exactly once, and they provide a complete characterization of the transition time $T$.

Returning to the characterization of $\chi(t)$: The first initial incumbent to transition to technology $B$ has initial knowledge ratio equal to $\chi(0)$, so we can define $\chi(t)$ to coincide with this

[37] Here $g_B(t) < g_B^*$ follows by noting that at each time $t < \infty$, fewer than $N$ firms innovate for $B$, and the distribution of relative qualities $H_B(z_B, t)$ is first-order stochastically dominated by the BGP distribution $H_B^*(z_B)$.

firm's knowledge ratio $\frac{k_B(t)}{k_A(t)}$ until it reaches the right side of the smooth-pasting condition (B4), after which $\chi$ remains constant. This ratio $\frac{k_B(t)}{k_A(t)}$ satisfies

$$\begin{aligned}\frac{d}{dt}\log\left(\frac{k_B(t)}{k_A(t)}\right) &= \frac{\sigma_I \dot{K}_B(t)}{k_B(t)} - \frac{\sigma_I \dot{K}_A(t)}{k_A(t)} - \lambda\eta_A s \\ &= \sigma_I \frac{\dot{K}_B(t) - \frac{k_B(t)}{k_A(t)}\dot{K}_A(t)}{\lambda K_B(0) + \sigma_I K_B(t)} - \lambda\eta_A s.\end{aligned}$$

Hence $\chi$ satisfies

$$\frac{\dot{\chi}(t)}{\chi(t)} = \left(\sigma_I \frac{\dot{K}_B(t) - \chi(t)\dot{K}_A(t)}{\lambda\sigma_E K_B(0) + \sigma_I K_B(t)} - \lambda\eta_A s\right)\mathbb{1}\left[\chi(t) \geq \frac{\eta_A}{\eta_B}\frac{\rho + \delta - \lambda\eta_B s}{\rho + \delta}\right]. \tag{B5}$$

Moreover, the arguments above imply that the cutoff $\chi(t)$ yields a valid description of firm innovation decisions: For any firm innovating for technology $A$ at $t = 0$, the ratio $\frac{k_B(t)}{k_A(t)}$ is strictly above $\chi(t)$ until it satisfies the smooth-pasting condition (B4), at which time the firm transitions to $B$.

**Step 3:** $K(t)$ **evolution.** To conclude the proof, I provide a self-contained version of the evolution equations (B2, B3) for the knowledge stocks. Given $\chi(0)$, let $z_{A0} \geq 1$ denote the initial relative quality of the firm with initial knowledge ratio $\frac{k_B(0)}{k_A(0)}$ equal to $\chi(0)$:

$$\frac{\lambda\sigma_E + \sigma_I}{\lambda\sigma_E z_{A0} + \sigma_I}\frac{K_B(0)}{K_A(0)} = \chi(0). \tag{B6}$$

Let $T_0$ denote the time at which this firm transitions to technology $B$. For $t \in [0, T_0)$, the knowledge stocks evolve according to

$$\begin{aligned}\ddot{K}_A(t) &= -[\delta - (\lambda + \sigma_I N_A(t))\,\eta_A s]\,\dot{K}_A(t), \\ \ddot{K}_B(t) &= -[\delta - (\lambda + \sigma_I N_B(t))\,\eta_B s]\,\dot{K}_B(t) + \delta N\,(\lambda\sigma_E + \sigma_I)\,\eta_B s K_B(t),\end{aligned}$$

where $N_A(t) = \exp(-\delta t)[N - H_{A0}(z_{A0})]$ and $N_B(t) = N - N_A(t)$. The initial conditions for these differential equations are

$$
\begin{aligned}
&K_A(0), \quad K_B(0), \\
&\frac{\dot{K}_A(0)}{K_A(0)} = \int_{z_{A0}}^{\infty} [\lambda \sigma_E z_A + \sigma_I] \eta_A s dH_{A0}(z_A), \\
&\frac{\dot{K}_B(0)}{K_B(0)} = [\lambda \sigma_E + \sigma_I] \eta_B s H_{A0}(z_{A0}).
\end{aligned} \tag{B7}
$$

To determine the time $T_0$, for arbitrary $z_A \geq z_{A0}$ let $k_A(t; z_A)$ solve the differential equation

$$
\dot{k}_A(t; z_A) = \lambda \eta_A s k_A(t; z_A) + \sigma_I \dot{K}_A(t)
$$

with initial condition $k_A(0; z_A) = (\lambda \sigma_E z_A + \sigma_I) K_A(0)$. Then setting $z_A = z_{A0}$, if $\lambda \eta_A s < g_B^*$ the time $T_0$ is the unique time at which the ratio

$$
\frac{\lambda \sigma_E K_B(0) + \sigma_I K_B(t)}{k_A(t; z_{A0})}
$$

is increasing and satisfies the smooth-pasting condition (B1). The inequality $t \geq T_0$ then holds if and only if

$$
\begin{aligned}
&\frac{\lambda \sigma_E K_B(0) + \sigma_I K_B(t)}{k_A(t; z_{A0})} \geq \frac{\eta_A}{\eta_B} \frac{\rho + \delta - \lambda \eta_B s}{\rho + \delta} \quad \text{and} \\
&\frac{\sigma_I \dot{K}_B(t)}{\lambda \sigma_E K_B(0) + \sigma_I K_B(t)} \geq \lambda \eta_A s + \frac{\sigma_I \dot{K}_A(t)}{k_A(t; z_{A0})}.
\end{aligned} \tag{B8}
$$

If $\lambda \eta_A s > g_B^*$, the second inequality in this condition cannot be satisfied; hence $T_0 = \infty$.

For $t \geq T_0$, we can explicitly identify the initial relative quality $z_A(0)$ of firms transitioning to technology $B$. This function is simply $\tilde{z}_A(t)$ defined above, and for $t \geq T_0$ it satisfies the smooth-pasting condition

$$
\frac{\lambda \sigma_E K_B(0) + \sigma_I K_B(t)}{k_A(t; \tilde{z}_A(t))} = \frac{\eta_A}{\eta_B} \frac{\rho + \delta - \lambda \eta_B s}{\rho + \delta}.
$$

Integrating the evolution equation for $k_A(t; z_A)$ reveals that this function is linear in $z_A$:

$$
k_A(t; z_A) = k_A(t; z_{A0}) + (z_A - z_{A0}) \exp(\lambda \eta_A s t) \lambda \sigma_E K_A(0).
$$

Substituting into the previous equation yields an implicit equation for $\tilde{z}_A(t)$:

$$\frac{\lambda\sigma_E K_B(0)+\sigma_I K_B(t)}{k_A(t;z_{A0})+(\tilde{z}_A(t)-z_{A0})\exp(\lambda\eta_A st)\,\lambda\sigma_E K_A(0)}=\frac{\eta_A}{\eta_B}\frac{\rho+\delta-\lambda\eta_B s}{\rho+\delta}.$$

Solving for $\tilde{z}_A(t)$ and recalling the characterization (B8) of $T_0$, we can write

$$\begin{aligned}&\tilde{z}_A(t) \\ &=z_{A0}+\frac{\exp(-\lambda\eta_A st)}{\lambda\sigma_E K_A(0)}\left\{[\lambda\sigma_E K_B(0)+\sigma_I K_B(t)]\frac{\eta_B}{\eta_A}\frac{\rho+\delta}{\rho+\delta-\lambda\eta_B s}-k_A(t;z_{A0})\right\}\mathbb{1}[(\text{B8})].\end{aligned} \tag{B9}$$

To summarize, I find that the knowledge stocks satisfy the dynamical system

$$\begin{cases}\ddot{K}_A(t) & =-[\delta-(\lambda+\sigma_I N_A(t))\eta_A s]\dot{K}_A(t) \\ & \quad -\frac{\rho+\delta}{\rho+\delta-\lambda\eta_B s}[\lambda\sigma_E K_B(0)+\sigma_I K_B(t)]\eta_B s\exp(-\delta t)h_{A0}(\tilde{z}_A(t))\dot{\tilde{z}}_A(t), \\ \ddot{K}_B(t) & =-[\delta-(\lambda+\sigma_I N_B(t))\eta_B s]\dot{K}_B(t)+\delta N(\lambda\sigma_E+\sigma_I)\eta_B sK_B(t) \\ & \quad +[\lambda\sigma_E K_B(0)+\sigma_I K_B(t)]\eta_B s\exp(-\delta t)h_{A0}(\tilde{z}_A(t))\dot{\tilde{z}}_A(t), \\ \dot{k}_A(t;z_{A0}) & =\lambda\eta_A sk_A(t;z_{A0})+\sigma_I\dot{K}_A(t),\end{cases} \tag{B10}$$

where $\tilde{z}_A$ is given by (B9) and $z_{A0}$ is given by (B6). The initial conditions are (B7) and $k_A(T_E;z_{A0})=(\lambda\sigma_E z_{A0}+\sigma_I)K_A(0)$.

**Step 4: $H_{A0}$ comparative static.** The comparative static with respect to $H_{A0}$ is monotone: If $\sigma_I=0$, firm innovation decisions are independent of all other firms after entry, so the cutoff $\chi(t)$ is invariant to $H_{A0}$.[38] If $\sigma_I>0$, initial incumbents are more inclined to start innovating for technology $A$ when the trajectory $K_A(t)$ is larger and the trajectory $K_B(t)$ is lower. With innovation decisions held fixed, a first-order stochastically increasing shift in the distribution $H_{A0}$ implies a larger growth rate for $K_A$ at fixed innovation decisions, and hence a larger level since $K_A(0)$ is fixed. This induces more incumbents to innovate for technology $A$, producing an upward shift in the trajectory of $K_A(t)$ and a downward shift in the trajectory of $K_B(t)$. These effects are mutually reinforcing, so $\chi(t)$ must be strictly increasing in $H_{A0}$. ■

### B.2.2 Equilibrium Converging to $A$

The following proposition characterizes all monotone equilibria converging to the BGP for technology $A$, analogous to Proposition 3 in Section 4.2:

[38] See also the explicit solution for this case in Section 4.3 or Appendix C.

**Proposition B.1.** *In any monotone equilibrium converging to the BGP for technology A, there exists a cutoff $\chi(t)$ such that a firm innovates for B if and only if*

$$\frac{k_B(t)}{k_A(t)} \geq \chi(t).$$

*There exists a time $T_E \geq 0$ such that $\chi(t) = \min\{\chi(0), \kappa(t)\}$ for $t \in [0, T_E]$, with $\chi(T_E) = \kappa(T_E)$. For $t \in (T_E, \infty)$, the cutoff solves the differential equation*

$$\frac{\dot{\chi}(t)}{\chi(t)} = \left(\lambda \eta_B s + \sigma_I \frac{\frac{1}{\chi(t)}\dot{K}_B(t) - \dot{K}_A(t)}{\lambda \sigma_E K_A(T_E) + \sigma_I K_A(t)}\right) \mathbb{1}\left[\chi(t) \leq \frac{\eta_A}{\eta_B} \frac{\rho + \delta}{\rho + \delta - \lambda \eta_A s}\right]. \tag{B11}$$

*The equilibrium is unique up to the parameters $\chi(0)$ and $T_E \geq 0$. The knowledge stocks $K(t)$ are the solutions to the dynamical systems (B19, B21).*

Note several differences from Proposition 3: First, firms choose $B$ provided that the weak inequality $\frac{k_B(t)}{k_A(t)} \geq \chi(t)$ holds, instead of the strong inequality $\frac{k_B(t)}{k_A(t)} > \chi(t)$. This ensures that entrants begin innovating for technology $B$ for $t \leq T_E$. The evolution equation (B11) ensures that $\chi(t)$ equals the knowledge ratio $\frac{k_B(t)}{k_A(t)}$ for the last entrant to innovate first for technology $B$, until the following smooth-pasting condition is satisfied:

$$\frac{k_B(t)}{k_A(t)} = \frac{\eta_A}{\eta_B} \frac{\rho + \delta}{\rho + \delta - \lambda \eta_A s}. \tag{B12}$$

After this time, $\chi(t)$ remains equal to the expression on the right side.

**Sketch of Proposition B.1.** I only show how to derive the dynamical system for the knowledge stocks $K(t)$; the remaining parts of the proof are analogous to that for Proposition 3.

Note first that at each time $t > T_E$, there are potentially two groups of firms that began innovating for technology $B$ but are now transitioning back to technology $A$: (i) initial incumbents with $z_A(0) < z_{A0}$, where $z_{A0}$ satisfies

$$\frac{\lambda \sigma_E + \sigma_I}{\lambda \sigma_E z_{A0} + \sigma_I} \frac{K_B(0)}{K_A(0)}; \tag{B13}$$

and (ii) entrants at each time $t_0 \in [0, T_E)$. We can track the transitions by initial incumbents just as in equilibria converging to technology $B$. Let $k_B(t)$ solve the differential equation

$$\dot{k}_B(t) = \lambda \eta_B s k_B(t) + \sigma_I \dot{K}_B(t),$$

with initial condition $k_B(0) = (\lambda \sigma_E + \sigma_I) K_B(0)$. The initial incumbent with $z_A = z_{A0}$ must be the first to transition back to technology $A$ (if at all). If $\lambda\eta_B s < g_A^*$, this transition time is the unique time at which the ratio

$$\frac{k_B(t)}{\lambda z_{A0}\sigma_E K_A(0) + \sigma_I K_A(t)}$$

is decreasing and satisfies the smooth-pasting condition (B12). Hence this time has passed if and only if

$$\begin{aligned} \frac{k_B(t)}{\lambda z_{A0}\sigma_E K_A(0) + \sigma_I K_A(t)} &\leq \frac{\eta_A}{\eta_B}\frac{\rho+\delta}{\rho+\delta-\lambda\eta_A s} \quad \text{and} \\ \lambda\eta_B s + \frac{\sigma_I \dot{K}_B(t)}{k_B(t)} &\leq \frac{\sigma_I \dot{K}_A(t)}{\lambda z_{A0}\sigma_E K_A(0) + \sigma_I K_A(t)}. \end{aligned} \tag{B14}$$

For each time $t$ after these conditions are satisfied, let $\tilde{z}_A(t)$ denote the initial relative quality of the incumbent transitioning back to technology $A$. Then $\tilde{z}_A(t)$ must solve the smooth-pasting condition

$$\frac{k_B(t)}{\lambda\sigma_E\tilde{z}_A(t)K_A(0) + \sigma_I K_A(t)} = \frac{\eta_A}{\eta_B}\frac{\rho+\delta}{\rho+\delta-\lambda\eta_A s}.$$

Setting $\tilde{z}(t) = z_{A0}$ before incumbents begin transitioning back to $A$, we can solve to find

$$\tilde{z}_A(t) = \begin{cases} z_{A0} & \text{if (B14) not satisfied,} \\ \frac{1}{\lambda\sigma_E K_A(0)}\left[\frac{\eta_B}{\eta_A}\frac{\rho+\delta-\lambda\eta_A s}{\rho+\delta}k_B(t) - \sigma_I K_A(t)\right] & \text{if (B14) satisfied.} \end{cases} \tag{B15}$$

Now consider entrants at time $t_0 \in [0, T_E)$ that potentially transition back to technology $A$. Let $k_B(t;t_0)$ solve the differential equation

$$\dot{k}_B(t;t_0) = \lambda\eta_B s k_B(t;t_0) + \sigma_I \dot{K}_B(t),$$

with initial condition $k_B(t;t_0) = (\lambda\sigma_E + \sigma_I)K_B(t_0)$. The entrant with $t_0 = T_E$ must be the first to transition back (if at all). If $\lambda\eta_B s < g_A^*$, this is the unique time at which the ratio

$$\frac{k_B(t;T_E)}{\lambda\sigma_E K_A(T_E) + \sigma_I K_B(t)}$$

is decreasing and satisfies the smooth-pasting condition (B12). Let $\bar{T}$ denote this time. We have $t \geq \bar{T}$ if and only if

$$\frac{k_B(t;T_E)}{\lambda\sigma_E K_A(T_E)+\sigma_I K_A(t)} \leq \frac{\eta_A}{\eta_B}\frac{\rho+\delta}{\rho+\delta-\lambda\eta_A s} \quad \text{and} \qquad \text{(B16)}$$
$$\lambda\eta_B s+\frac{\sigma_I \dot{K}_B(t)}{k_B(t;T_E)} \leq \frac{\sigma_I \dot{K}_A(t)}{\lambda\sigma_E K_A(T_E)+\sigma_I K_A(t)}.$$

For $t \geq \bar{T}$, let $\tilde{t}_0(t)$ denote the entry time of the firms transitioning back to technology $A$. Then $\tilde{t}_0(t)$ must solve the smooth-pasting condition

$$\frac{k_B(t;\tilde{t}_0(t))}{\lambda\sigma_E K_A(\tilde{t}_0(t))+\sigma_I K_A(t)} = \frac{\eta_A}{\eta_B}\frac{\rho+\delta}{\rho+\delta-\lambda\eta_A s}. \qquad \text{(B17)}$$

To derive a differential equation for $\tilde{t}_0(t)$, note first that we can integrate the evolution equation for $k_B(t;t_0)$ to write

$$k_B(t;t_0)=\exp\left(\lambda\eta_B s(t-t_0)\right)(\lambda\sigma_E+\sigma_I)K_B(t_0)+\int_{t_0}^{t}\exp\left(\lambda\eta_B s(t-\tau)\right)\sigma_I\dot{K}_B(\tau)d\tau.$$

Differentiating in $t_0$ yields

$$\frac{\partial k_B(t;t_0)}{\partial t_0}=\exp\left(\lambda\eta_B s(t-t_0)\right)\left[-\lambda\eta_B s\left(\lambda\sigma_E+\sigma_I\right)K_B(t_0)+\lambda\sigma_E\dot{K}_B(t_0)\right].$$

We can then differentiate the smooth-pasting condition (B17) to find that for $t \geq \bar{T}$,

$$\dot{\tilde{t}}_0(t) \qquad \text{(B18)}$$
$$=\frac{\lambda\eta_B s\frac{\eta_A}{\eta_B}\frac{\rho+\delta}{\rho+\delta-\lambda\eta_A s}\left[\lambda\sigma_E K_A(\tilde{t}_0(t))+\sigma_I K_A(t)\right]+\sigma_I\dot{K}_B(t)-\frac{\eta_A}{\eta_B}\frac{\rho+\delta}{\rho+\delta-\lambda\eta_A s}\sigma_I\dot{K}_A(t)}{\frac{\eta_A}{\eta_B}\frac{\rho+\delta}{\rho+\delta-\lambda\eta_A s}\lambda\sigma_E\dot{K}_A(\tilde{t}_0(t))-\exp\left(\lambda\eta_B s(t-\tilde{t}_0(t))\right)\left[\lambda\sigma_E\dot{K}_B(\tilde{t}_0(t))-\lambda\eta_B s\left(\lambda\sigma_E+\sigma_I\right)K_B(\tilde{t}_0(t))\right]}.$$

By the same calculation as in the proof of Proposition 3, for $t<\bar{T}$ the knowledge stocks satisfy the dynamical system

$$\begin{cases} \ddot{K}_A(t) & = -[\delta - (\lambda + \sigma_I N_A(t))\,\eta_A s]\,\dot{K}_A(t) + \delta N(\lambda\sigma_E + \sigma_I)\eta_A s K_A(t)\mathbb{1}[t \geq T_E] \\ & \quad + \frac{\rho+\delta-\lambda\eta_A s}{\rho+\delta} k_B(t)\eta_B s \exp(-\delta t)\, h_A^*(\tilde{z}_A(t))\left(-\dot{\tilde{z}}_A(t)\right), \\ \ddot{K}_B(t) & = -[\delta - (\lambda + \sigma_I N_B(t))\,\eta_B s]\,\dot{K}_B(t) + \delta N(\lambda\sigma_E + \sigma_I)\eta_B s K_B(t)\mathbb{1}[t < T_E] \\ & \quad - k_B(t)\eta_B s \exp(-\delta t)\, h_A^*(\tilde{z}_A(t))\left(-\dot{\tilde{z}}_A(t)\right), \\ \dot{k}_B(t) & = \lambda\eta_B s k_B(t) + \sigma_I \dot{K}_B(t), \\ \dot{N}_A(t) & = \delta\,[N\mathbb{1}[t \geq T_E] - N_A(t)] + \exp(-\delta t)\, h_A^*(\tilde{z}_A(t))\left(-\dot{\tilde{z}}_A(t)\right), \end{cases} \tag{B19}$$

where $\tilde{z}_A$ is given by (B15) and $z_{A0}$ is given by (B13). The initial conditions are

$$\begin{aligned} & K_A(0), \quad K_B(0), \\ & \frac{\dot{K}_A(0)}{K_A(0)} = \int_{z_{A0}}^{\infty} [\lambda\sigma_E z_A + \sigma_I]\,\eta_A s\, dH_A^*(z_A), \\ & \frac{\dot{K}_B(0)}{K_B(0)} = [\lambda\sigma_E + \sigma_I]\,\eta_B s H_A^*(z_{A0}), \end{aligned} \tag{B20}$$

as well as $k_B(0) = (\lambda\sigma_E + \sigma_I)K_B(0)$ and $N_A(0) = N - H_A^*(z_{A0})$.

Solving the dynamical system (B19) forward, the time $\bar{T}$ is the first time that (B16) is satisfied. For $t \geq \bar{T}$, the knowledge stocks instead satisfy

$$\begin{cases} \ddot{K}_A(t) & = -[\delta - (\lambda + \sigma_I N_A(t))\,\eta_A s]\,\dot{K}_A(t) + \delta N(\lambda\sigma_E + \sigma_I)\eta_A s K_A(t) \\ & \quad + \frac{\rho+\delta-\lambda\eta_A s}{\rho+\delta} k_B(t)\eta_B s \exp(-\delta t)\, h_A^*(\tilde{z}_A(t))\left(-\dot{\tilde{z}}_A(t)\right) \\ & \quad + [\lambda\sigma_E K_A(\tilde{t}_0(t)) + \sigma_I K_A(t)]\,\eta_A s \exp(-\delta(t - \tilde{t}_0(t)))\,\delta N\left(-\dot{\tilde{t}}_0(t)\right), \\ \ddot{K}_B(t) & = -[\delta - (\lambda + \sigma_I N_B(t))\,\eta_B s]\,\dot{K}_B(t) \\ & \quad - k_B(t)\eta_B s \exp(-\delta t)\, h_A^*(\tilde{z}_A(t))\left(-\dot{\tilde{z}}_A(t)\right) \\ & \quad - \frac{\rho+\delta}{\rho+\delta-\lambda\eta_A s} [\lambda\sigma_E K_A(\tilde{t}_0(t)) + \sigma_I K_A(t)]\,\eta_A s \exp(-\delta(t - \tilde{t}_0(t)))\,\delta N\left(-\dot{\tilde{t}}_0(t)\right), \\ \dot{k}_B(t) & = \lambda\eta_B s k_B(t) + \sigma_I \dot{K}_B(t), \\ \dot{N}_A(t) & = \delta\,[N - N_A(t)] + \exp(-\delta t)\, h_A^*(\tilde{z}_A(t))\left(-\dot{\tilde{z}}_A(t)\right) \\ & \quad + \exp(-\delta(t - \tilde{t}_0(t)))\,\delta N\left(-\dot{\tilde{t}}_0(t)\right), \\ \dot{\tilde{t}}_0(t) & = \text{(B18)}, \end{cases} \tag{B21}$$

where the initial condition for $\tilde{t}_0$ is $\tilde{t}_0(\bar{T}) = T_E$. ∎

## B.3 Equilibrium Existence: Proof of Proposition 4

In this section, I provide a full proof of the existence of monotone equilibria converging to the BGP for technology $B$, given the assumptions of Proposition 4. I sketch the analogous proof for equilibria converging to the BGP for $A$, which is almost identical. Finally, I show that multiple equilibria arise whenever $\lambda, \sigma_I > 0$, and I prove comparative statics for the thresholds $\kappa_A^*$ and $\kappa_B^*$ in the initial distribution $H_{A0}$.

### B.3.1 Equilibrium Converging to $B$

Each monotone equilibrium is uniquely determined by the initial cutoff $\chi(0) \geq \frac{K_B(0)}{K_A(0)}$. I let $z_{A0}$ denote the initial relative quality for an initial incumbent at the cutoff:

$$\frac{\lambda\sigma_E + \sigma_I}{\lambda\sigma_E z_{A0} + \sigma_I}\kappa(0) = \chi(0).$$

The following proposition proves the existence of a monotone equilibrium converging to the BGP for technology $B$ under weaker conditions than assumed in Proposition 4:

**Proposition B.2.** *There exists a threshold $\kappa_B^*$ such that a monotone equilibrium converging to the BGP for technology $B$ exists if the following hold:*

*(i) For each $t \geq 0$*

$$\lambda\eta_B s + \frac{\sigma_I \dot{K}_B(t)}{k_B(t)} \geq \frac{\sigma_I \dot{K}_A(t)}{k_A(t)}, \tag{B22}$$

*where $\dot{k}_A(t) = \sigma_I \dot{K}_A(t)$ and $\dot{k}_B(t) = \lambda\eta_B s k_B(t) + \sigma_I \dot{K}_B(t)$ with $k_\theta(0) = (\lambda\sigma_E + \sigma_I)K_\theta(0)$;*

*(ii) $\kappa(0) \geq \kappa_B^*$.*

*The initial cutoff $\chi(0)$ is such that $z_{A0}$ is a stable solution to the fixed-point equation (B23). The threshold $\kappa_B^*$ is strictly increasing in $H_{A0}$ when $\lambda > 0$.*

**Proof of Proposition B.2:** Fix an initial cutoff $\chi(0)$ and the corresponding initial relative quality $z_{A0}$. The purpose of the proof is to demonstrate that the innovation decisions implied by the resulting cutoff $\chi(t)$ are privately optimal for all firms.

First consider an initial incumbent with relative quality $z_A \geq z_{A0}$. This firm initially innovates for technology $A$, and the arguments in Step 3 of the proof of Proposition 3 demonstrate that the innovation decisions implied by the cutoff $\chi(t)$ are optimal.

Now consider an initial incumbent with relative quality $z_A < z_{A0}$. This firm initially innovates for technology $B$, and the innovation decisions implied by the cutoff require that it never innovates for technology $A$. The firm's research productivities then evolve according to

$$\frac{\dot{k}_A(t)}{k_A(t)} = \frac{\sigma_I \dot{K}_A(t)}{k_A(t)} \quad \text{and} \quad \frac{\dot{k}_B(t)}{k_B(t)} = \lambda \eta_B s + \frac{\sigma_I \dot{K}_B(t)}{k_B(t)}.$$

If the incumbent reverted back to technology $A$ at some time $T$, Lemma B.2 would require that $k_A$ and $k_B$ satisfy the smooth-pasting condition

$$k_B(T)\eta_B \Psi_B(T) = k_A(T)\eta_A \Psi_A(T),$$

with the corresponding second-order necessary condition

$$\lambda \eta_B s + \frac{\sigma_I \dot{K}_B(t)}{k_B(t)} + \rho + \delta - \frac{1}{\Psi_B(T)} \leq \frac{\sigma_I \dot{K}_A(t)}{k_A(t)} + \rho + \delta - \lambda \eta_A s - \frac{1}{\Psi_A(T)}.$$

Given the bounds $\Psi_\theta(T) \in \left[\frac{1}{\rho+\delta}, \frac{1}{\rho+\delta-\lambda\eta_\theta s}\right]$, this inequality is ruled out precisely by (B22).

Finally, note that entrants at $t_0 \geq 0$ innovate for technology $B$ permanently if and only if this yields higher value than innovating initially for technology $A$ and transitioning to technology $B$ at a later stopping time $T(t_0) \geq t_0$. This stopping time is either $\infty$ (if $\sigma_I = 0$ or $\lambda \eta_A s \geq g_B^*$) or is given by the unique solution to the smooth-pasting condition

$$\frac{k_B(T(t_0); t_0)}{k_A(T(t_0); t_0)} = \frac{\eta_A}{\eta_B} \frac{\rho + \delta - \lambda \eta_B s}{\rho + \delta},$$

where $k_A$ and $k_B$ satisfy the evolution equations

$$\frac{\dot{k}_A(t; t_0)}{k_A(t; t_0)} = \lambda \eta_A s + \frac{\sigma_I \dot{K}_A(t)}{k_A(t; t_0)} \quad \text{and} \quad \frac{\dot{k}_B(t; t_0)}{k_B(t; t_0)} = \frac{\sigma_I \dot{K}_B(t)}{k_B(t; t_0)},$$

with initial conditions $k_\theta(t_0; t_0) = (\lambda \sigma_E + \sigma_I) K_\theta(t_0)$.

If an entrant at $t_0$ permanently innovates for technology $B$, let $q_\theta^B(t; t_0)$ denote the corresponding quality for technology $\theta$ at time $t$. These qualities satisfy the evolution equations

$$\begin{aligned}
\dot{q}_A^B(t; t_0) &= 0 \\
\dot{q}_B^B(t; t_0) &= \left[\lambda q_B^B(t; t_0) + \sigma_I K_B(t)\right] \eta_B s,
\end{aligned}$$

with initial conditions $q_\theta^B(t_0, t_0) = \sigma_E K_\theta(t_0)$. If an entrant instead innovates initially for technology $A$, let $q_\theta^A(t; t_0)$ denote the corresponding quality for technology $\theta$. The evolution equations are now

$$\dot{q}_A^A(t;t_0) = \mathbb{1}_{t<T(t_0)}\left[\lambda q_A^A(t;t_0) + \sigma_I K_A(t)\right]\eta_A s$$
$$\dot{q}_B^B(t;t_0) = \mathbb{1}_{t\geq T(t_0)}\left[\lambda q_B^B(t;t_0) + \sigma_I K_B(t)\right]\eta_B s,$$

with the same initial conditions as above. The entrant's value when choosing to innovate technology $\theta$ initially is

$$V_E^\theta(t_0) \equiv \bar{\pi}\int_{t_0}^{\infty} \exp\left(-(\rho+\delta)(t-t_0)\right)\left[q_A^\theta(t;t_0) + q_B^\theta(t;t_0)\right]dt.$$

Entrants at $t_0$ innovate for $B$ if and only if $V_E^B(t_0) \geq V_E^A(t_0)$.

To determine when this condition holds, note first that the dynamical system (B10) that describes the evolution of the knowledge stocks is linearly homogeneous in $K(t)$ conditional on $z_{A0}$. As a result, there exists a function $\tilde{K}(t;z_{A0},\kappa(0))$ that depends on initial conditions only through $z_{A0}$ and $\kappa(0)$ such that

$$K(t) = \tilde{K}(t;z_{A0},\kappa(0))K_A(0).$$

This normalization implies $\tilde{K}_A(0;z_{A0},\kappa(0)) = 1$ and $\tilde{K}_B(0;z_{A0},\kappa(0)) = \kappa(0)$. To understand how these normalized knowledge stocks depend on $z_{A0}$ and $\kappa(0)$, hold $\kappa(0)$ fixed and consider an increase in $z_{A0}$, or equivalently a decrease in $\chi(0)$. This adjustment raises the initial growth rate of $\tilde{K}_B$ and lowers the initial growth rate of $\tilde{K}_A$, implying that initial incumbents transition more rapidly to $B$. These effects are mutually reinforcing, and they imply that $\tilde{K}_A$ and $\tilde{K}_B$ are respectively strictly decreasing and strictly increasing in $z_{A0}$ at $t > 0$. Now hold $z_{A0}$ fixed, and consider an increase in $\kappa(0)$. This leaves the initial growth rates of $\tilde{K}_A$ and $\tilde{K}_B$ unaffected, and if all initial incumbents with $z_A > z_{A0}$ never transition to $B$, it only raises $\tilde{K}_B$ for $t \geq 0$ while leaving $\tilde{K}_A$ unchanged. But if initial incumbents eventually transition to $B$, the increase in $\tilde{K}_B$ induces the incumbents to transition more rapidly. This leads to a further increase in $\tilde{K}_B$ and a decrease in $\tilde{K}_A$ for $t$ sufficiently large.

These linear homogeneity and comparative dynamics observations carry over directly to the qualities $q_\theta^{\theta'}(t;t_0)$ and the values $V_E^\theta(t_0)$. In particular, the difference

$$\frac{V_E^B(t_0)}{K_A(0)} - \frac{V_E^A(t_0)}{K_A(0)}$$

depends on the initial conditions $K(0)$ only through $\kappa(0)$, and the Envelope Theorem implies that it is strictly increasing in $\kappa(0)$ and $z_{A0}$ when $T(t_0) > 0$. With $g_B(t) \uparrow 1$ and $g_A(t) \downarrow 0$, there also exists a finite time $t_0$ after which the difference is (weakly) positive. The Intermediate Value Theorem implies a function $\kappa^E(t, z_{A0})$ such that $V_E^B(t_0) \geq V_E^A(t_0)$ if and only if

$$\kappa(0) \geq \kappa^E(t_0, z_{A0}),$$

where $\kappa^E$ is strictly decreasing in $z_{A0}$ and strictly decreasing in $t_0$ for $t_0$ sufficiently large.[39]

Now consider optimality for initial incumbents. Given an incumbent with relative quality $z_A$, denote the value of innovating initially for technology $\theta$ by

$$\tilde{V}^\theta(z_A; z_{A0}, \kappa(0))K_A(0),$$

where the same arguments as above imply linear homogeneity in $[K(t)]_t$. Similarly, the difference

$$\Delta(z_A; z_{A0}, \kappa(0)) \equiv \tilde{V}^B(z_A; z_{A0}, \kappa(0)) - \tilde{V}^A(z_A; z_{A0}, \kappa(0))$$

is also strictly increasing in $z_{A0}$ and $\kappa(0)$ when the firm's transition time is positive, while the Envelope Theorem implies that it is strictly decreasing in $z_A$ (again when the transition time is positive). In equilibrium, $z_{A0} \geq 1$ must be such that this difference is zero:[40]

$$\Delta(z_{A0}; z_{A0}, \kappa(0)) = 0. \tag{B23}$$

Note first that a solution can only exist for $\kappa(0)$ sufficiently large: (B23) is trivially violated when $\kappa(0) = 0$. Fixing any $z_A \geq 1$, we have that $\Delta(z_A; z_A, \kappa(0)) \to \infty$ as $\kappa(0) \to \infty$, so

[39] In the case with $\sigma_I = 0$ or $\lambda\eta_A s \geq g_B$, we can derive an explicit formula:

$$\kappa^E(t, z_{A0}) = \frac{\frac{\lambda\eta_A s}{\rho+\delta-\lambda\eta_A s}\frac{\sigma_E \tilde{K}_A(t; z_{A0})}{\rho+\delta} + \int_t^\infty \exp\left(-(\rho+\delta)(\tau-t)\right)\frac{\eta_A s \sigma_I \tilde{K}_A(\tau; z_{A0})}{\rho+\delta-\lambda\eta_A s}d\tau}{\frac{\lambda\eta_B s}{\rho+\delta-\lambda\eta_B s}\frac{\sigma_E \tilde{K}_B(t; z_{A0})}{\rho+\delta} + \int_t^\infty \exp\left(-(\rho+\delta)(\tau-t)\right)\frac{\eta_B s \sigma_I \tilde{K}_B(t; z_{A0})}{\rho+\delta-\lambda\eta_B s}d\tau}.$$

Here $\tilde{K}_A(t; z_{A0}) \equiv K_A(t)/K_A(0)$ as in the text of the proof, while for this expression I instead define $\tilde{K}_B(t; z_{A0}) \equiv K_B(t)/K_B(0)$. I can normalize both knowledge stocks independently because the function $\tilde{z}_A(t)$ described in the dynamical system (B10) is constant in this case, so the differential equations for $K$ are not interdependent conditional on $z_{A0}$.

[40] Using the same assumptions and notation as in footnote 39, this equation can be written explicitly:

$$\frac{\lambda\eta_A s}{\rho+\delta-\lambda\eta_A s}\frac{\sigma_E}{\rho+\delta}z_{A0} = \left[\frac{\lambda\eta_B s}{\rho+\delta-\lambda\eta_B s}\frac{\sigma_E}{\rho+\delta} + \int_0^\infty \exp\left(-(\rho+\delta)\tau\right)\frac{\eta_B s \sigma_I \tilde{K}_B(\tau; z_{A0})}{\rho+\delta-\lambda\eta_B s}d\tau\right]\kappa(0)$$
$$- \int_0^\infty \exp\left(-(\rho+\delta)\tau\right)\frac{\eta_A s \sigma_I \tilde{K}_A(\tau; z_{A0})}{\rho+\delta-\lambda\eta_A s}d\tau.$$

we can take any $\kappa(0)$ such that $\Delta(z_A; z_A, \kappa(0)) > 0$. It is also straightforward to see that $\Delta(z_A; z_A, \kappa(0)) \to -\infty$ as $z_A \to \infty$: Even with all initial incumbents innovating for technology $B$, an incumbent's value from doing so remains uniformly bounded. By the Intermediate Value Theorem, these arguments imply that for $\kappa(0)$ sufficiently large, there exists a solution $z_{A0}$ to the fixed-point equation (B23). This solution must also be *stable* in the sense that $\Delta(z_A; z_A, \kappa(0))$ is strictly declining in $z_A$ in a neighborhood of the solution $z_{A0}$. The Implicit Function Theorem then guarantees that this solution is strictly increasing in $\kappa(0)$.

To conclude, recall that entrant optimality requires $\kappa(0) \geq \kappa^E(t_0, z_{A0})$ for all $t_0 \geq 0$. With $z_{A0} \geq 1$ and $\kappa(0) > 0$ fixed, there exists a time $\bar{t} \geq 0$ such that this inequality is satisfied for $t_0 \geq \bar{t}$. With $\kappa^E$ strictly decreasing in $z_{A0}$ and $z_{A0}$ strictly increasing in $\kappa(0)$ for any stable solution to (B23), we also have that the entry condition can is satisfied for all $t_0 \geq 0$ provided that $\kappa(0)$ is sufficiently large. ■

The following lemma verifies that the condition (B22) is implied by the assumptions of Proposition 4.

**Lemma B.2.** *If $\lambda\eta_B s \geq \frac{\sigma_I}{\lambda\sigma_E+\sigma_I} g_A^*$, then (B22) holds in any equilibrium with $\ddot{K}_A(t) \leq 0$ for all $t \geq 0$ and $g_A(0) \leq g_A^*$.*

**Proof of Lemma B.2:** Differentiating the right side of (B22) in $t$ yields

$$\frac{\partial}{\partial t} \frac{\sigma_I \dot{K}_A(t)}{k_A(t)} = \frac{\sigma_I \ddot{K}_A(t)}{k_A(t)} - \left(\frac{\sigma_I \dot{K}_A(t)}{k_A(t)}\right)^2 \leq 0.$$

Hence the right side of (B22) is bounded above by

$$\frac{\sigma_I K_A(0)}{k_A(0)} g_A(0) = \frac{\sigma_I}{\lambda\sigma_E + \sigma_I} g_A(0) \leq \frac{\sigma_I}{\lambda\sigma_E + \sigma_I} g_A^*.$$

■

### B.3.2 Equilibrium Converging to $A$

The following proposition states the existence result for equilibria converging to the BGP for technology $A$, analogous to Proposition B.2:

**Proposition B.3.** *There exists a threshold $\kappa_A^* \geq \kappa_B^*$ such that a monotone equilibrium converging to the BGP for technology $A$ exists if $\kappa(0) \leq \kappa_A^*$.*

Note the absence of a condition on the growth rates of the knowledge stocks $K$. In any monotone equilibrium converging to $A$, at some time $T_E$ entrants must start by innovating for technology $A$. This only holds if the knowledge stock $K_A$ always grows relative to the knowledge stock $K_B$. As a result, any firm that initially innovates for technology $A$ never finds it optimal to begin innovating for technology $B$.

**Proof of Proposition B.3:** Many proof details are similar to those of Proposition B.2, so I only highlight the distinctions. As Proposition B.1 indicates, a monotone equilibrium converging to the BGP for technology $A$ is fully characterized by $z_{A0} \geq 1$ and $T_E \geq 0$: All initial incumbents with relative qualities $z_A(0) \geq z_{A0}$ permanently innovate for technology $A$, while those with $z_A(0) < z_{A0}$ initially innovate for technology $B$. Similarly, all entrants at time $t_0 \geq T_E$ permanently innovate for technology $A$, while entrants at time $t_0 < T_E$ initially innovate for technology $B$. An equilibrium consists of a pair $(z_{A0}, T_E)$ such that these decisions are optimal.

For each $z_{A0} \geq 0$, define the function $T_E(z_{A0})$ such that $T_E(z_{A0}) \in [0, \infty]$ gives the smallest time at which the entrant decisions described above are optimal. To see that such a time exists, suppose first $T_E = 0$. If entrants at time $t_0 = 0$ find it optimal to innovate initially for $A$, then so must every subsequent entrant given that $K_A$ grows relative to $K_B$. We can then set $T_E(z_{A0}) = 0$. Otherwise, we can raise $T_E$ until we find a time $T_E(z_{A0}) > 0$ at which entrants are indifferent to their initial innovation directions. Necessarily each entrant before $T_E(z_{A0})$ must strictly prefer to innovate initially for $B$, while each subsequent entrant strictly prefers to innovate for $A$. If such an indifference time does not exist, then we can set $T_E(z_{A0}) = \infty$, and the economy instead converges asymptotically to the BGP for technology $B$. With positive spillovers to entrants and (potentially) across incumbents, we must have $T_E(z_{A0})$ weakly increasing in $z_{A0}$, and strictly so whenever $T_E(z_{A0})$ is interior.

Given the function $T_E(z_{A0})$, we must now find an equilibrium relative quality cutoff $z_{A0}$. First set $z_{A0} = 1$. If, given $T_E(1)$, an initial incumbent with relative quality $z_A(0) = 1$ prefers to innovate for technology $A$, then we must have $T_E(1) = 0$. We have then found a "corner" equilibrium with $z_{A0} = 1$ and no innovation for technology $B$. Otherwise, we can raise $z_{A0}$ until we find a value at which an initial incumbent with relative quality $z_{A0}$ is indifferent its initial innovation direction. Necessarily each initial incumbent with $z_A(0) < z_{A0}$ strictly prefers to innovate initially for $B$, while each initial incumbent with $z_A(0) > z_{A0}$ strictly prefers to innovate for technology $A$. To see that such a value $z_{A0}$ exists, note first that for $z_A(0)$ sufficiently high an initial incumbent strictly prefers to innovate initially for technology $A$ regardless of other firms' innovation decisions. But we cannot immediately apply the Intermediate Value Theorem: There exists a minimum value $\underline{z} \in [1, \infty]$ such that $T_E(z) = \infty$ for $z \geq \underline{z}$. If $\underline{z} < \infty$ and $\sigma_I > 0$, then each incumbent's net value from innovating initially for technology $B$ jumps upward at $z_{A0} = \underline{z}$. This holds because a positive mass of future firms have altered their innovation directions, and with $\sigma_I > 0$ initial incumbents find it valuable to align with them. But this poses no issues for the argument: If the indifference cutoff $z_{A0}$ appears before $\underline{z}$, then we have found a monotone equilibrium converging to the BGP for technology $A$. If the indifference cutoff $z_{A0}$ appears after (or at) $\underline{z}$, then the Intermediate Value Theorem implies a monotone equilibrium converging to the BGP for technology $B$.

This argument establishes the existence of a monotone equilibrium regardless of the initial knowledge stock ratio $\kappa(0)$. But clearly $T_E(z_{A0})$ is strictly increasing in $\kappa(0)$: By the same argument as in the proof of Proposition B.2, entrants always have greater incentives to start innovating for technology $B$ when its initial knowledge stock $K_B(0)$ increases relative to technology $A$'s. Hence the minimum value $\underline{z}$ defined above is decreasing in $\kappa(0)$. Since clearly $T_E(z_{A0}) = 0$ when $\kappa(0) = 0$, there must exist a threshold $\kappa_A^*$ such that the indifference cutoff $z_{A0}$ is reached before $\underline{z}$ if and only if $\kappa(0) \leq \kappa_A^*$. This provides the threshold characterization for the existence of a monotone equilibrium converging to $A$ stated in the Proposition.

Finally, since we have established that a monotone equilibrium always exists, we must have $\kappa_A^* \geq \kappa_B^*$. ■

### B.3.3 Equilibrium Multiplicity

It remains to show that the economy can feature multiple monotone equilibria when $\lambda > 0$ and $\sigma_I > 0$. To see this, fix $\kappa(0) = \kappa_B^*$, and let $z_{A0}^*$ denote the corresponding equilibrium cutoff for incumbent relative qualities. Note that we must have $z_{A0}^* > 1$ to sustain a transition to technology $B$.

Using the function $T_E(z_{A0})$ defined in the proof of Proposition B.3, observe that we must have $T_E(z_{A0}) < \infty$ for any $z_{A0} < z_{A0}^*$. Otherwise, we could reduce both $\kappa(0)$ and $z_{A0}^*$ to maintain a monotone equilibrium that converges to the BGP for technology $B$, a contradiction to the definition of $\kappa_B^*$. Hence $T_E(z_{A0})$ jumps down discretely as $z_{A0}$ is reduced from $z_{A0}^*$. But with $\sigma_I > 0$, this yields a jump upward in each incumbent's net value from innovating initially for technology $A$ instead of technology $B$. Since the marginal incumbent with $z_A(0) = z_{A0}^*$ was before indifferent, the incumbent with $z_A(0) = z_{A0}^* - \varepsilon$ must strictly prefer technology $A$ for $\varepsilon > 0$ sufficiently small. Hence by decreasing $z_{A0}$ from $z_{A0}^*$ we can find a new indifference point (or a corner solution with $z_{A0} = 1$) that yields a monotone equilibrium converging to the BGP for technology $A$.

This argument establishes multiplicity of equilibria at $\kappa(0) = \kappa_B^*$. But it applies almost unchanged to $\kappa(0) = \kappa_B^* + \varepsilon$ for $\varepsilon > 0$ small, because $z_{A0}^*$ is continuous in $\kappa(0)$ for $\kappa(0) > \kappa_B^*$. They key observation is again that $\sigma_I > 0$ implies a discrete jump in incentives at $z_{A0} = \underline{z}$. Hence $\kappa_B^* < \kappa_A^*$ when $\sigma_I > 0$. ■

# C Equilibrium with Technology Choice: Benchmarks

In this appendix, I provide a full characterization of the economy's equilibria in two benchmark cases. In Appendix C.1, firms build exclusively on the aggregate knowledge stock when innovating ($\lambda = 0$), reflecting the most common way of modeling innovation in existing work on directed innovation and technological transitions. I show that this assumption leaves no role for firm heterogeneity or the concentraton of R&D to affect the economy's equilibrium. In Appendix C.2, I address the other extreme in which firms build exclusively on their own past advances ($\sigma_I = 0$), so that all knowledge spillovers are confined to entrants. As discussed in Section 4.3, in this case the economy's equilibrium depends richly on the initial distribution of incumbents and the concentration of scientists. Both cases feature a unique equilibrium, which is also monotone in the sense of Definition 3.

## C.1 External Knowledge Accumulation: $\lambda = 0$

### C.1.1 Equilibrium

Suppose firms build exclusively on external knowledge when innovating ($\lambda = 0 < \sigma_I$), and consider the problem of a firm with qualities $q(t_0)$ at time $t_0 \geq 0$ choosing its innovation direction at each time $t \geq t_0$. We can directly integrate the quality evolution equation (8) to find

$$q_\theta(t) = q_\theta(t_0) + \sigma_I \eta_\theta \int_{t_0}^{t} K_\theta(\tau) s_\theta(q(\tau), \tau) d\tau.$$

The firm's objective (11) can then be written

$$\frac{\bar{\pi}(q_A(t_0) + q_B(t_0))}{\rho + \delta} + \bar{\pi} \int_{t_0}^{\infty} \exp\left(-(\rho + \delta)(t - t_0)\right) \sum_{\theta \in \{A,B\}} \frac{\sigma_I K_\theta(t) \eta_\theta s_\theta(q(t), t)}{\rho + \delta} dt.$$

The first term gives the discounted value of profits given the initial qualities $q(t_0)$, and the second term incorporates the additional value generated by innovation at each time $t \geq t_0$. This objective function is linear in the allocation of scientists at each time $t \geq t_0$, so the firm's solution is simply to allocate all scientists toward the technology $\theta$ with the largest marginal productivity of research $K_\theta(t)\eta_\theta$ at each time $t \geq t_0$. Equivalently, a firm exclusively innovates for technology $B$ at time $t$ if and only if the knowledge ratio satisfies $\kappa(t) \geq \eta_A/\eta_B$, and otherwise exclusively innovating for technology $A$. Comparing to the more general analysis of Section 4.1, here we also obtain a corner solution for the allocation of scientists because each technology's innovation rate is linear in the mass of scientists. This case additionally implies that each firm's allocation is identical, because $\lambda = 0$ and flow profits are linear in qualities.

This threshold characterization of firm innovation decisions carries over to the economy's equilibrium. When $\kappa(t) \geq \eta_A/\eta_B$, all innovation is directed toward technology $B$, so that the knowledge stock for $B$ grows relative to that for $A$. The ratio $\kappa(t)$ is then strictly increasing, so that innovation is permanently directed toward $B$. The opposite holds when $\kappa(t) < \eta_A/\eta_B$. Thus the ratio of research productivities $\eta_A/\eta_B$ functions as a threshold for $\kappa(0)$ that determines the equilibrium direction of technological change:

**Proposition C.1.** *The economy with $\lambda = 0$ has a unique equilibrium. All firms exclusively innovate for technology $B$ at $t \geq 0$ if and only if*

$$\kappa(0) \geq \frac{\eta_A}{\eta_B},$$

*with all innovation directed toward A otherwise. If innovation is directed toward technology B, the knowledge stock $K_B$ immediately grows at the BGP rate $g_B^* = \sigma_I \eta_B S$, and $Q_B(t)/K_B(t)$ increases monotonically to its BGP value.*

**Proof of Proposition C.1:** The threshold characterization of equilibrium research follows immediately from the discussion preceding the proposition. If research is directed toward technology $B$, the evolution equation (9) for $K_B$ immediately implies $\dot{K}_B(t) = \sigma_I \eta_B S K_B(t)$. Integrating the evolution equation (15) for $Q_B$ then yields

$$Q_B(t) = \frac{g_B + \delta N \sigma_E}{g_B + \delta} K_B(t) - \frac{(1 - N\sigma_E)\, g_B}{g_B + \delta} K_B(0) \exp(-\delta t).$$

Hence

$$\frac{Q_B(t)}{K_B(t)} = \frac{g_B + \delta N \sigma_E}{g_B + \delta} - \frac{(1 - N\sigma_E)\, g_B}{g_B + \delta} \exp(-(\delta + g_B)t)$$

By (19), this ratio is strictly increasing in $t$ and limits to the BGP value from Proposition 1. ■

This result is a special case of Propositions 3 and 4. The economy's unique equilibrium features a transition to technology $B$ if and only if the initial ratio of knowledge stocks $\kappa(0)$ is sufficiently high. But with $\eta_B > \eta_A$, the economy may transition even if technology $B$ is initially inferior to technology $A$ ($\kappa(0) < 1$). Along a transition, the knowledge stock $K_B$ features no transitional dynamics, while the aggregate quality $Q_B$ increases relative to $K_B$ as incumbents improve relative to entrants before exit.

### C.1.2 Efficiency

A key implication of this analysis is that the equilibrium direction of technological change is essentially myopic: At each time, all firms research the technology with the higher research productivity $\sigma_I K_\theta(t)\eta_\theta$. This holds because firms do not internalize the value of their knowledge spillovers for future innovating firms, and with $\lambda = 0$ they do not benefit directly from their past knowledge production. To understand the resulting inefficiency, consider the problem of a social planner who can choose each firm's allocation of scientists to maximize the consumer's discounted utility (1), but cannot otherwise modify the equilibrium. This social planner solves

$$\max_{[s_\theta(q,t)]_{\theta,q,t}} \int_0^\infty \exp(-\rho t) C(t) dt, \tag{C1}$$

subject to the resource constraint $s_A(q,t)+s_B(q,t) \leq s$ and with all remaining quantities determined in equilibrium. The following proposition solves this problem for the case with $\lambda = 0$, demonstrating that the equilibrium transitions to technology $B$ insufficiently often:

**Proposition C.2.** *With $\lambda = 0$, the social planner allocates all scientists to technology $B$ at $t \geq 0$ if and only if*

$$\kappa(0) \geq \frac{\eta_A}{\eta_B} J,$$

*with all innovation directed toward $A$ otherwise. The adjustment factor $J \in (0,1)$ is independent of initial conditions.*

**Proof of Proposition C.2:** Since incumbents' qualities do not affect their research productivities, it is optimal for the social planner to permanently allocate all scientists to one technology. If the social planner chooses technology $A$, then $Q_A$ continues to grow at the BGP rate $g_A^* = \sigma_I \eta_A S$, while $Q_B$ remains constant at its initial value $Q_B(0) = N\sigma_E K_B(0)$. With flow consumption equal to $\bar{C}(Q_A + Q_B)$ for a constant $\bar{C} > 0$, this yields social value

$$\begin{aligned}
\frac{U^A}{\bar{C}} &= \frac{Q_A(0)}{\rho - g_A^*} + \frac{Q_B(0)}{\rho} \\
&= \frac{1}{\rho - g_A^*} \frac{g_A^* + \delta N \sigma_E}{g_A^* + \delta} K_A(0) + \frac{1}{\rho} N \sigma_E K_B(0).
\end{aligned}$$

If instead the social planner permanently allocates all scientists to technology $B$, then $Q_A$ and $Q_B$ evolve according to

$$\dot{Q}_A(t) = \delta\left[N\sigma_E K_A(0) - Q_A(t)\right],$$
$$\dot{Q}_B(t) = g_B^* K_B(0)\exp\left(g_B^* t\right) + \delta\left[N\sigma_E K_B(0)\exp\left(g_B^* t\right) - Q_B(t)\right].$$

In the second equation I make use of the relation $K_B(t) = K_B(0)\exp\left(g_B^* t\right)$. Integrating these equations yields

$$Q_A(t) = N\sigma_E K_A(0) + \frac{(1-N\sigma_E)\,g_A^*}{g_A^*+\delta}K_A(0)\exp(-\delta t),$$
$$Q_B(t) = \frac{g_B^*+\delta N\sigma_E}{g_B^*+\delta}K_B(0)\exp\left(g_B^* t\right) - \frac{(1-N\sigma_E)\,g_B^*}{g_B^*+\delta}K_B(0)\exp(-\delta t).$$

The social welfare from researching technology $B$ permanently is then

$$\begin{aligned}\frac{U^B}{\bar{C}} &= \int_0^\infty \exp(-\rho t)\left[Q_A(t)+Q_B(t)\right]dt\\ &= \left[\frac{1}{\rho}N\sigma_E + \frac{1}{\rho+\delta}\frac{(1-N\sigma_E)\,g_A^*}{g_A^*+\delta}\right]K_A(0)\\ &\quad + \left[\frac{1}{\rho-g_B^*}\frac{g_B^*+\delta N\sigma_E}{g_B^*+\delta} - \frac{1}{\rho+\delta}\frac{(1-N\sigma_E)\,g_B^*}{g_B^*+\delta}\right]K_B(0).\end{aligned}$$

Researching technology $B$ is socially optimal if and only if $U^B \geq U^A$, or equivalently

$$j(g_B^*)g_B^*K_B(0) \geq j(g_A^*)g_A^*K_A(0),$$

where I define

$$j(g) \equiv \left[\frac{1}{\rho-g}\frac{g+\delta N\sigma_E}{g+\delta} - \frac{1}{\rho+\delta}\frac{(1-N\sigma_E)\,g}{g+\delta} - \frac{1}{\rho}N\sigma_E\right]\frac{1}{g}.$$

This function can equivalently be written

$$j(g) = \frac{\frac{g+\delta N\sigma_E}{\rho-g} + \frac{\delta}{\rho+\delta}(1-N\sigma_E)}{g+\delta}\frac{1}{\rho}.$$

Direct calculation implies that this function is strictly increasing in $g$:

$$j'(g) \propto (g+\delta)\left(1+\frac{g+\delta N\sigma_E}{\rho-g}\right) - \left[g+\delta N\sigma_E + (\rho-g)\frac{\delta}{\rho+\delta}(1-N\sigma_E)\right].$$

This value is positive if and only if

$$\delta(1-N\sigma_E)+(g+\delta)\frac{g+\delta N\sigma_E}{\rho-g}>-(\rho-g)\frac{\delta}{\rho+\delta}(1-N\sigma_E).$$

This necessarily holds, because the left side is positive and the right side is negative because of the assumed upper bound on entrant spillovers (19) and the assumed lower bound on the discount rate (20). This analysis demonstrates that the social planner exclusively and permanently researches technology $B$ if and only if

$$\kappa(0)\geq\frac{\eta_A}{\eta_B}J,$$

where the adjustment factor $J\equiv j(g_A^*)/j(g_B^*)$ is positive but strictly smaller than one. ■

For intuition, consider the limit $\sigma_E N\uparrow 1$, so that the aggregate qualities $Q$ are unaffected by entry and exit and hence display no transitional dynamics. The factor $J$ then limits to

$$\lim_{\sigma_E N\uparrow 1} J=\frac{\rho-g_B^*}{\rho-g_A^*}.$$

This value is strictly smaller than one precisely because the *long-run* growth rate for technology $A$ is smaller than that for technology $B$. The social planner internalizes the value of knowledge spillovers for future innovation, so her transition threshold takes into account not only the $t=0$ research productivities for each technology, but also the implied long-run growth rates $g_A^*$ and $g_B^*$.

Finally, note the distinction between Proposition C.2 and Proposition 6, which is the corresponding efficiency result for the benchmark case with only internal knowledge ($\sigma_I=0$). With only external knowledge, the social planner prefers to transition to technology $B$ more often than in equilibrium; but conditional on the direction of innovation, the equilibrium is efficient (Proposition C.2). With only internal knowledge, the social planner does not necessary prefer to transition to technology $B$ more often; but conditional on the direction of innovation, the equilibrium converges too slowly to the limiting BGP.

## C.2 Internal Knowledge Accumulation: $\sigma_I=0$

### C.2.1 Equilibrium

Now consider the other extreme case in which firms build exclusively on internal knowledge when innovating ($\sigma_I=0<\lambda$), summarized in Section 4.3. With $\sigma_I=0$, we can again integrate the quality evolution equation (8) to find

$$q_\theta(t) = q_\theta(t_0)\exp\left(\lambda\eta_\theta \int_{t_0}^{\tau} s_\theta(q(\tau),\tau)d\tau\right).$$

In contrast to the case with $\lambda = 0$, $q_\theta(t)$ is not additively separable between the initial quality $q_\theta(t_0)$ and the allocation of scientists. Since firms build on their past advances, an increase in the initial quality $q_\theta(t_0)$ raises the firm's research productivity for $\theta$ at each time $\tau \geq t_0$. With profits linear in quality, this mechanism leads to path dependence whereby a firm is more likely to research a technology that it has researched previously. The following lemma precisely characterizes equilibrium innovation decisions:

**Lemma C.1.** *With $\sigma_I = 0$, a firm with qualities $q$ exclusively innovates for technology $B$ if*

$$\frac{\eta_B}{\rho+\delta-\lambda\eta_B s} q_B \geq \frac{\eta_A}{\rho+\delta-\lambda\eta_A s} q_A,$$

*with all innovation directed toward $A$ otherwise.*

**Proof of Lemma C.1:** It is clearly optimal for the firm to innovate permanently for one technology. If the firm innovates permanently for technology $A$, its value is

$$\frac{V^A}{\bar{\pi}} = \frac{q_A}{\rho+\delta-\lambda\eta_A s} + \frac{q_B}{\rho+\delta}$$

If the instead innovates permanently for technology $B$, its value is

$$\frac{V^B}{\bar{\pi}} = \frac{q_A}{\rho+\delta} + \frac{q_B}{\rho+\delta-\lambda\eta_B s}.$$

Innovating for technology $B$ is optimal if and only if $V^B \geq V^A$. Substituting the expressions above into this inequality yields the inequality stated in the Lemma. This inequality is self-reinforcing, so an incumbent's innovation direction is perfectly persistent. ∎

This result clarifies the two forces that determine the direction of a firm's innovation. First, as discussed above, a firm has a greater propensity to innovate for the technology for which it has a higher quality. Second, firms have a greater propensity to innovate for technology $B$ because of its higher basic research productivity $\eta_B > \eta_A$. This force is stronger with $\lambda > 0$ because the firm's problem is genuinely foward-looking: Innovating for technology $\theta$ raises the quality $q_\theta$, making research for $\theta$ more productive in the future. The firm internalizes this dynamic effect because its innovation decision has a non-negligible impact on the change in quality $q_\theta$ (in contrast to the knowledge stock $K_\theta$).

We can apply this result to derive a convenient characterization of entrant and incumbent innovation decisions. Substituting the entrant qualities $q_\theta^E(t) = \sigma_E K_\theta(t)$ into the inequality stated in the Lemma, we observe that an entrant at $t \geq 0$ permanently innovates for technology $B$ if and only if the knowledge stock ratio $\kappa(t)$ is larger than the *entry threshold*

$$\kappa^E \equiv \frac{\eta_A}{\rho+\delta-\lambda\eta_A s}\left(\frac{\eta_B}{\rho+\delta-\lambda\eta_B s}\right)^{-1} \in (0,1).$$

An incumbent at $t=0$ with initial qualities $q_A(0) = z_A(0) q_A^E(0)$ and $q_B(0) = q_B^E(0)$ permanently innovates for technology $B$ if and only if its relative quality $z_A(0)$ is above the cutoff

$$z_{A0}^* \equiv \frac{\kappa(0)}{\kappa^E}$$

Note that path dependence implies that $t=0$ incumbents always have a lower propensity to innovate for $B$ than entrants at $t=0$.

Using this characterization of innovation decisions, the following lemma derives a simple dynamical system for the knowledge stocks $K$, analogous to that of Proposition 3:

**Lemma C.2.** *With $\sigma_I = 0$, the knowledge stocks $K$ evolve according to*

$$\ddot{K}_A(t) = -(\delta-\lambda\eta_A s)\dot{K}_A(t) + \delta N \lambda \eta_A s \sigma_E K_A(t) \mathbb{1}\left[\kappa(t) < \kappa^E\right], \tag{C2}$$

$$\ddot{K}_B(t) = -(\delta-\lambda\eta_B s)\dot{K}_B(t) + \delta N \lambda \eta_B s \sigma_E K_B(t) \mathbb{1}\left[\kappa(t) \geq \kappa^E\right]. \tag{C3}$$

*The corresponding initial conditions are*

$$\begin{aligned} &K_A(0), \quad K_B(0), \\ &\frac{\dot{K}_A(0)}{K_A(0)} = \lambda\eta_A s \sigma_E \int_{z_{A0}^*}^{\infty} z_A dH_A(z_A), \\ &\frac{\dot{K}_B(0)}{K_B(0)} = \lambda\eta_B s \sigma_E H_A(z_{A0}^*). \end{aligned} \tag{C4}$$

**Proof of Lemma C.2:** For expositional purposes only, suppose $\kappa(0) > \kappa^E$, and fix $t$ small enough that $\kappa(t') > \kappa^E$ for $t' \in [0,t]$. Consider technology $A$. Using Lemma C.1 and integrating the evolution equation (8) for $q_A$, we have that for any $t=0$ incumbent with $z_A(0) > z_{A0}^*$,

$$q_A(t) = z_A(0)\sigma_E K_A(0) \exp(\lambda\eta_A s t).$$

The density of these incumbents at $t=0$ is $h_A(z_A(0))$. But with exit at rate $\delta>0$, the density at time $t>0$ falls to $\exp(-\delta t)h_A(z_A(0))$. Hence the evolution equation (9) for $K_A$ can be written

$$\dot{K}_A(t)=\lambda\eta_A s\sigma_E K_A(0)\left(\int_{z_A>z^*_{A0}} z_A dH_A(z_A)\right)\exp(-(\delta-\lambda\eta_A s)t).$$

Differentiating in $t$ yields (C2).

Turning to technology $B$, the same argument as above implies that for any $t=0$ incumbent with $z_A(0)\leq z^*_{A0}$,

$$q_B(t)=\sigma_E K_B(0)\exp(\lambda\eta_B st).$$

By time $t$, the total mass of $t=0$ incumbents with $z_A(0)\leq z^*_{A0}$ is $\exp(-\delta t)H_A(\hat{z}_A)$. Similarly, for any firm who entered at time $\tau\in[0,t]$,

$$q_B(t)=\sigma_E K_B(\tau)\exp(\lambda\eta_B s(t-\tau)).$$

By time $t\geq\tau$, the density of these firms is $\exp(-\delta(t-\tau))\delta N$. Hence the evolution equation (9) can be written

$$\begin{aligned}\dot{K}_B(t)=&\lambda\eta_B s\sigma_E K_B(0)H_A(z^*_{A0})\exp(-(\delta-\lambda\eta_B s)t)\\&+\lambda\eta_B s\sigma_E\delta N\int_0^t K_B(\tau)\exp(-(\delta-\lambda\eta_B s)(t-\tau))d\tau.\end{aligned}$$

Differentiating in $t$ yields (C3).

The same arguments yield the evolution equations (C2, C3) regardless of $\kappa(0)$ or $t$. The initial conditions (C4) follow directly from the evolution equation (9) for the knowledge stocks $K$, given the assumption that the economy is following the BGP for technology $A$ at $t=0$. ∎

The benchmark case with $\sigma_I=0$ is convenient precisely because the system (C2, C3) is autonomous and can be integrated in closed form. The following lemma demonstrates this for the case when the economy transitions to the BGP for technology B:

**Lemma C.3.** *Suppose entrants innovate for technology B at each time, $\kappa(t)\geq\kappa^E$ for $t\geq0$.*

*(i) The solution to the system (C2, C3) is*

$$\begin{aligned}\frac{K_A(t)}{K_A(0)}&=c_{A1}-c_{A2}\exp(-(\delta-\lambda\eta_A s)t),\\\frac{K_B(t)}{K_B(0)}&=c_{B1}\exp\left(-\left(g^*_B+\delta-\lambda\eta_B s\right)t\right)+c_{B2}\exp\left(g^*_B t\right),\end{aligned}$$

*where $c_{A1}, c_{A2}, c_{B1}, c_{B2} > 0$ satisfy the initial conditions (C4).*

*(ii) The knowledge ratio $\kappa(t) = K_B(t)/K_A(t)$ is strictly increasing and strictly convex in the initial condition $\kappa(0)$ for $t > 0$, and there exists a time $\hat{t} \geq 0$ such that $(t - \hat{t})\dot{\kappa}(t) > 0$.*

**Proof of Lemma C.3:** Integrating the evolution equation (C2) for $\dot{K}_A$ yields

$$\frac{K_A(t)}{K_A(0)} = 1 + \lambda\eta_A s\sigma_E \left(\int_{z_{A0}^*}^{\infty} z_A dH_A(z_A)\right) \frac{1 - \exp(-(\delta - \lambda\eta_A s)\,t)}{\delta - \lambda\eta_A s}.$$

Hence the integration constants from the Lemma are

$$\begin{aligned} c_{A1} &= 1 + c_{A2}, \\ c_{A2} &= \lambda\eta_A s\sigma_E \left(\int_{z_{A0}^*}^{\infty} z_A dH_A(z_A)\right) \frac{1}{\delta - \lambda\eta_A s}. \end{aligned}$$

Integrating the evolution equation (C3) for $\dot{K}_B$ and making use of the formula for $g_B^*$ given in Proposition 1 yields the expression for $K_B(t)/K_B(t)$ given in the Lemma statement. The integration constants jointly satisfy the initial conditions

$$\begin{aligned} 1 &= c_{B1} + c_{B2}, \\ \lambda\eta_B s\sigma_E H_A(z_{A0}^*) &= -\left(g_B^* + \delta - \lambda\eta_B s\right) c_{B1} + g_B^* c_{B2}. \end{aligned}$$

The solution to this system is

$$\begin{aligned} c_{B1} &= \frac{g_B^* - \lambda\eta_B s\sigma_E H_A(z_{A0}^*)}{2g_B^* + \delta - \lambda\eta_B s}, \\ c_{B2} &= \frac{g_B^* + \delta - \lambda\eta_B s\left(1 - \sigma_E H_A(z_{A0}^*)\right)}{2g_B^* + \delta - \lambda\eta_B s}. \end{aligned}$$

Clearly $c_{B2} > 0$ given the assumed lower bound (14) on the exit rate $\delta$. We also have

$$\begin{aligned} g_B^* - \lambda\eta_B s\sigma_E H_A(z_{A0}^*) &\geq g_B^* - \lambda\eta_B S\sigma_E \\ &\geq g_B^* - \lambda\eta_B s \\ &= -\frac{\delta - \lambda\eta_B s}{2} - \lambda\eta_B S\sigma_E + \sqrt{\left(\frac{\delta - \lambda\eta_B s}{2}\right)^2 + \lambda\sigma_E\delta\eta_B S}. \end{aligned}$$

Direct calculation implies that the right side is positive given the assumed upper bound on spillovers to entrants (19), so $c_{B1} > 0$.

To prove the properties of $\kappa$ stated in the Lemma, note that the above analysis shows that $K_A(t)/K_A(0)$ depends on the initial knowledge stocks $K(0)$ only through $z^*_{A0}$. This cutoff is strictly increasing in $\kappa(0)$ while $K_A(t)/K_A(0)$ is strictly decreasing in $\hat{z}_A$ for $t > 0$, so $K_A(t)/K_A(0)$ is strictly decreasing in $\kappa(0)$ for $t > 0$. A similar argument implies that $K_B(t)/K_B(0)$ is strictly increasing in $\kappa(0)$ for $t > 0$. Hence

$$\kappa(t) = \kappa(0)\frac{K_B(t)/K_B(0)}{K_A(t)/K_A(0)}$$

depends on $K(0)$ only through $\kappa(0)$, and it is strictly increasing and strictly convex in $\kappa(0)$ for $t > 0$. Finally, note that the growth rate of $\kappa$ satisfies

$$\frac{\dot{\kappa}(t)}{\kappa(t)} = \frac{\dot{K}_B(t)}{K_B(t)} - \frac{\dot{K}_A(t)}{K_A(t)}.$$

We can directly calculate

$$\begin{aligned}
\frac{\dot{K}_A(t)}{K_A(t)} &= \frac{(\delta - \lambda\eta_A s)\,c_{A2}}{c_{A1}\exp\left((\delta - \lambda\eta_A s)\,t\right) - c_{A2}}, \\
\frac{\dot{K}_B(t)}{K_B(t)} &= g^*_B\frac{-\left(1 + \frac{\delta - \lambda\eta_B s}{g^*_B}\right)c_{B1} + c_{B2}\exp\left(\left(2g^*_B + \delta - \lambda\eta_B s\right)t\right)}{c_{B1} + c_{B2}\exp\left(\left(2g^*_B + \delta - \lambda\eta_B s\right)t\right)}.
\end{aligned}$$

Since all integration coefficients are positive, we immediately have that the growth rate of $K_A(t)$ is declining over time, while the growth rate of $K_B(t)$ is increasing over time. The previous equation then implies that $\dot{\kappa}(t)$ is single-crossing from below, so there exists a time $\hat{t} \geq 0$ such that $\dot{\kappa}(t) < 0$ for $t < \hat{t}$ and $\dot{\kappa}(t) > 0$ for $t > \hat{t}$. ■

The second part of the lemma leverages the solution to the system (C2, C3) to prove two properties of the ratio $\kappa(t)$: It is strictly increasing in its initial value $\kappa(0)$, and it is generally "U-shaped" over time. These properties are useful because the solution in the lemma only describes the dynamics of the knowledge stocks $K(0)$ while entrants continue to research technology $B$, $\kappa(t) \geq \kappa^E$. If this condition is ever violated, the economy fails to transition to technology $B$ and instead converges back to the BGP for technology $A$. The properties of $\kappa(t)$ described in Lemma C.3 ensure that a transition takes place exactly when $\kappa(0)$ is sufficiently large.

**Proposition C.3.** *With $\sigma_I = 0$, the economy has a unique equilibrium. There exists a threshold $\kappa^* > \kappa^E$ such that if $\kappa(0) \geq \kappa^*$, all firms innovate for $B$ as $t \to \infty$, and the economy converges to the BGP for $B$. Otherwise, all firms innovate for $A$ as $t \to \infty$, and the economy converges to the*

*BGP for A. The economy displays transitional dynamics when $\kappa(0) > \kappa^E$.*

**Proof of Proposition C.3:** Clearly if $\kappa(0) \leq \kappa^E$, all innovation is initially directed toward technology $A$, so that $\kappa(t) < \kappa^E$ for all $t > 0$. Hence all incumbents innovate for technology $A$, so that the economy continues along the BGP for $A$.

If instead $\kappa(0) > \kappa^E$ but $\kappa(t) < \kappa^E$ for some time $t > 0$, then the economy again converges back to the BGP for $A$. To see this, let $\underline{t} = \inf\{t : \kappa(t) < \kappa^E\}$. For $t$ in a neighborhood to the right of $\underline{t}$, Lemma C.2 implies that the knowledge stocks $K$ evolve according to

$$\ddot{K}_A(t) = -(\delta - \lambda\eta_A s)\dot{K}_A(t) + \lambda\eta_A s\delta N\sigma_E K_A(t),$$
$$\ddot{K}_B(t) = -(\delta - \lambda\eta_B s)\dot{K}_B(t).$$

As in Lemma C.3, the solution implies that the growth rate $\dot{K}_A(t)/K_A(t)$ is increasing for $t$ in a neighborhood to the right of $\underline{t}$, while the growth rate $\dot{K}_B(t)/K_B(t)$ is decreasing. Since $\kappa$ must be strictly decreasing for $t$ in a neighborhood to the left of $\underline{t}$, this implies $\dot{K}_A(t)/K_A(t) > \dot{K}_B(t)/K_B(t)$ in a neighborhood to the right of $\underline{t}$. But then the inequality $\kappa(t) < \kappa^E$ is self-reinforcing, and the economy converges back to the BGP for technology $A$.

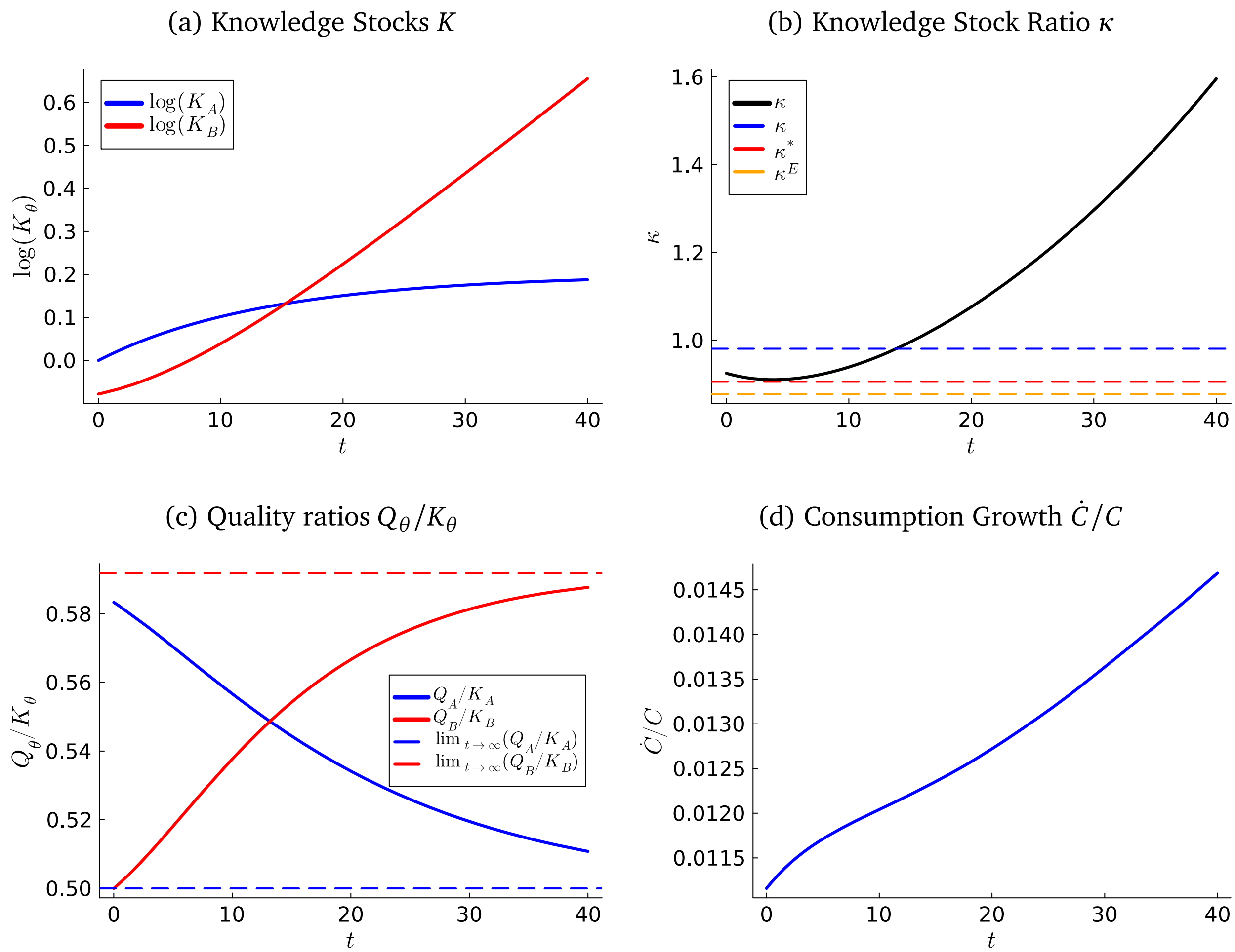
The argument above implies that the economy transitions to the BGP for technology $B$ asymptotically if and only if the trajectory of $\kappa(t)$ implied by (C2, C3) with initial conditions (C4) satisfies $\kappa(t) \geq \kappa^E$ for all $t \geq 0$. Lemma C.3 implies that $\kappa(t)$ is asymptotically increasing in $t$ and strictly increasing in $\kappa(0)$, so there must exist a threshold $\kappa^*$ such that $\kappa(t) \geq \kappa^E$ for all $t \geq 0$ if and only if $\kappa(0) \geq \kappa^*$. Note that this threshold must satisfy $\kappa^* > \kappa^E$, because otherwise $\kappa(t)$ initially falls below $\kappa^E$.

The transitional dynamics of the knowledge stocks $K$ and the aggregate qualities $Q$ are described by the system (15, C2, C3). This system is conveniently block diagonal, so that $K$ can be recovered without integrating the evolution equations for $Q$. ■

Figure C.1 displays the trajectories of the knowledge stocks $K$, the growth rates $\dot{K}_\theta/K_\theta$, the knowledge ratio $\kappa$, and the ratios $Q_\theta/K_\theta$ for an example transition. Note that the trajectory of the knowledge stock ratio $\kappa(t)$ is "U-shaped," initially decreasing before increasing asymptotically.

The threshold $\kappa^*$ determines the economy's propensity to transition in equilibrium, and it depends richly on model primitives. To gain intuition for the key forces, note that Lemma C.3 implies that a simple sufficient condition for the economy to transition is for the initial growth rate of $K_B$ to dominate the initial growth rate of $K_A$. Given the initial conditions (C4), this holds

Figure C.1: Example Transition with $\sigma_I = 0$




*Notes:* To calibrate, I set $\lambda = 1$, $\sigma_E = 0.5$, and $\sigma_I = 0$. I also set $\delta = 0.1$, $S = 1$, and $s = 1$, and I choose $\eta_A$ and $\eta_B$ to deliver BGP growth rates $g_A^* = 0.02$ and $g_B^* = 0.0225$. Finally, I set $\rho = 0.075$, and I specify the initial conditions $K_A(0) = 1$ and $K_B(0) = 0.925$ so that $\kappa(0) > \kappa^* \approx 0.91$.

if and only if

$$\eta_B H_A\left(z_{A0}^*\right) \geq \eta_A \int_{z_{A0}^*}^{\infty} z_A dH_A(z_A). \tag{C5}$$

This inequality depends on $\kappa(0)$ only through the relative quality cutoff $\hat{z}_A$, which is strictly increasing in $\kappa(0)$. As $\kappa(0)$ rises, the left side increases as a larger mass of $t = 0$ incumbents transition to $B$, raising the initial growth rate of $K_B$. The right side instead decreases as fewer $t = 0$ incumbents innovate for $A$, lowering the initial growth rate of $K_A$. There exists a unique value $\bar{\kappa}$ at which the inequality (C5) binds:

$$1 = \frac{\eta_A}{\eta_B} \frac{1}{1-\xi_A^*} \frac{\frac{\bar{\kappa}}{\kappa^E}}{\left(\frac{\bar{\kappa}}{\kappa^E}\right)^{1/\xi_A^*} - 1}. \tag{C6}$$

Here I have used the expression for $z_{A0}^*$ from Lemma C.1 and the BGP relative quality distribution from Proposition 1, which is a standard Pareto distribution when $\sigma_I = 0$.

The right side of equation (C6) is strictly increasing in $\xi_A^*$ and $\kappa^E$ and strictly decreasing in $\bar{\kappa}$, and it delivers a simple but powerful intuition about the drivers of a technological transition: Any change that thickens the tail of the old technology's firm-quality distribution slows the transition, because it raises both the relative mass of incumbents who choose not to transition and their initial innovation rates. Both effects increase their collective influence over the aggregate direction of innovation, which may be decisive if it induces entrants to switch back to innovating for technology $A$. However, any change that raises incentives for new firms to innovate for the new technology instead accelerates the transition by raising the relative mass of incumbents who choose to transition. The tail parameter $\xi_A^*$ and the entry threshold $\kappa^E$ respectively capture these two forces, but they depend on many of the same model primitives. The following proposition provides explicit comparative statics for $\bar{\kappa}$:

**Proposition C.4.** *With $\sigma_I = 0$, there exists a threshold $\bar{\kappa} \geq \kappa^*$ such that*

$$\frac{\dot{K}_B(0)}{K_B(0)} \geq \frac{\dot{K}_A(0)}{K_A(0)} \iff \kappa(0) \geq \bar{\kappa}.$$

*The threshold $\bar{\kappa}$ is strictly decreasing in $\sigma_E$, $S$, and $\eta_B$, and it is strictly increasing in $\rho$ and $\eta_A$. For each variable $v \in \{\delta, \lambda, s\}$, there exists a discount rate $\rho^v \geq 0$ that depends on model primitives such that*

*(i) $\bar{\kappa}$ is strictly increasing (decreasing) in $\delta$ locally if and only if $\rho$ is smaller (larger) than $\rho^\delta$;*

*(ii) $\bar{\kappa}$ is strictly increasing (decreasing) in $\lambda$ locally if and only if $\rho$ is larger (smaller) than $\rho^\lambda$;*

*(iii) $\bar{\kappa}$ is strictly increasing (decreasing) in $s$ locally if and only if $\rho$ is larger (smaller) than $\rho^s$.*

**Proof of Proposition C.4:** The existence of the threshold $\bar{\kappa} \geq \kappa^*$ follows immediately from the discussion preceding the Proposition.

Throughout the remainder of the proof, let RHS denote the right side of (C6), and let

$\nu \equiv \bar{\kappa}/\kappa^E$. It is immediate that RHS is strictly decreasing in $\nu$ and strictly increasing in $\xi_A^*$:

$$\frac{\partial \text{RHS}}{\partial \nu} = -\frac{\eta_A}{\eta_B}\frac{1}{1-\xi_A^*}\frac{\left(\frac{1}{\xi_A^*}-1\right)\nu^{1/\xi_A^*}+1}{\left(\nu^{1/\xi_A^*}-1\right)^2},$$
$$\frac{\partial \text{RHS}}{\partial \xi_A^*} = \text{RHS}\left[\frac{1}{1-\xi_A^*}+\frac{1}{(\xi_A^*)^2}\frac{\nu^{1/\xi_A^*}\log(\nu)}{\nu^{1/\xi_A^*}-1}\right]$$

Several comparative statics follow immediately: $\bar{\kappa}$ is strictly decreasing in $\sigma_E$ and $S$ because these parameters only reduce $\xi_A^*$. Similarly, $\bar{\kappa}$ is strictly increasing in $\rho$ because an increase in $\rho$ only increases $\kappa^E$. Now RHS is directly decreasing in $\eta_B$ and indirectly decreasing in $\eta_B$ through $\kappa^E$, so $\bar{\kappa}$ is strictly decreasing in $\eta_B$. A symmetric argument shows that $\bar{\kappa}$ is strictly increasing in $\eta_A$.

For the comparative static with respect to $v \in \{\delta, \lambda, s\}$, we can differentiate to find

$$\frac{d\text{RHS}}{dv} = \frac{\partial \text{RHS}}{\partial \xi_A^*}\frac{\partial \xi_A^*}{\partial v} + \frac{\partial \text{RHS}}{\partial \nu}\nu\kappa^E\frac{\partial (\kappa^E)^{-1}}{\partial v}. \tag{C7}$$

Now (C6) determines $\nu$ independently of $\kappa^E$, and since RHS depends on $\rho$ only through $\kappa^E$, this implies that $\nu$ is invariant to $\rho$. The first summand in (C7) depends on $\kappa^E$ only through $\nu$, and similarly for the partial derivative $\partial\text{RHS}/\partial\nu$, so these terms are invariant to $\rho$. With Lemma C.1, we can directly calculate

$$\frac{\partial (\kappa^E)^{-1}}{\partial \delta} = -(\kappa^E)^{-1}\left[\frac{1}{\rho+\delta-\lambda\eta_B s}-\frac{1}{\rho+\delta-\lambda\eta_A s}\right],$$
$$\frac{\partial (\kappa^E)^{-1}}{\partial \lambda} = (\kappa^E)^{-1}\left[\frac{\eta_B}{\rho+\delta-\lambda\eta_B s}-\frac{\eta_A}{\rho+\delta-\lambda\eta_A s}\right]s,$$
$$\frac{\partial (\kappa^E)^{-1}}{\partial s} = (\kappa^E)^{-1}\left[\frac{\eta_B}{\rho+\delta-\lambda\eta_B s}-\frac{\eta_A}{\rho+\delta-\lambda\eta_A s}\right]\lambda.$$

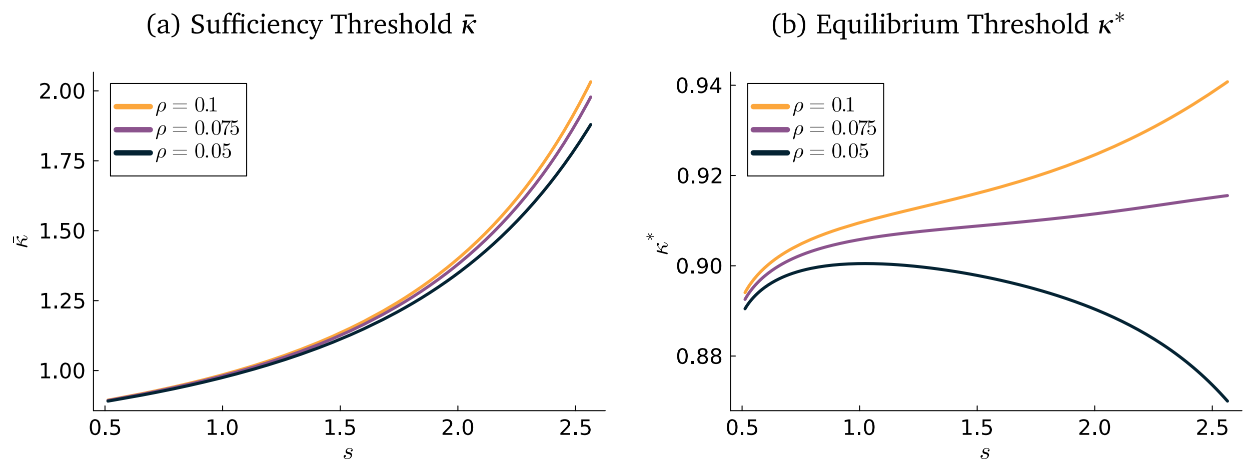
With $\sigma_I = 0$, Corollary 1 yields

$$\frac{\partial \xi_A^*}{\partial \delta} < 0 \quad \text{and} \quad \frac{\partial \xi_A^*}{\partial \lambda}, \frac{\partial \xi_A^*}{\partial s} > 0.$$

Hence (C7) implies that $d\text{RHS}/d\delta$ is strictly decreasing in $\rho$, becoming negative in the limit $\rho \to \infty$. There must then exist a value $\rho^\delta \geq 0$ such that $\bar{\kappa}$ is strictly increasing in $\delta$ locally if and only if $\rho < \rho^\delta$, while the opposite holds for $\rho > \rho^\delta$. The same argument implies the existence of the values $\rho^\lambda, \rho^s \geq 0$ and the corresponding comparative statics with respect to $\lambda$ and $\delta$ stated in the Proposition. ■

Comparative statics for the thresholds $\bar{\kappa}$ and $\kappa^*$ with respect to $s$ are illustrated in Figure

Figure C.2: Comparative Statics for Thresholds $\bar{\kappa}$ and $\kappa^*$




*Notes:* I vary the discount rate $\rho$ around its baseline value $\rho = 0.075$. The remaining parameters are exactly as in Figure C.1.

C.2.

### C.2.2 Efficiency

In the case with $\sigma_I = 0$, the solution to the planner's problem (C1) is complex: The planner generally chooses different innovation directions for different firms and may reverse these directions over time. To develop intuition about equilibrium inefficiencies, I consider the simpler problem in which the planner can choose the initial relative quality cutoff $z_{A0} \geq 1$ describing initial incumbents' innovation decisions. The planner then solves

$$\max_{z_{A0} \geq 1} \int_0^\infty \exp(-\rho t) C(t) dt, \tag{C8}$$

with entrant innovation decisions and all remaining quantities determined in equilibrium. The following proposition characterizes properties of the solution $\hat{z}_{A0}$:

**Proposition C.5.** *A solution $\hat{z}_{A0}$ to the social planner's problem (C8) exists and depends on $K(0)$ only through $\kappa(0)$. There exists a threshold $\hat{\kappa}$ such that*

*(i) the solution $\hat{z}_{A0}$ yields a transition to technology B if and only if $\kappa(0) \geq \hat{\kappa}$;*

*(ii) $\hat{z}_{A0} > z^*_{A0}$ if $\kappa(0) \geq \hat{\kappa}$; and*

*(iii) $\hat{z}_{A0} \leq z^*_{A0}$ if $\kappa(0) < \hat{\kappa}$, with equality only if $z^*_{A0} = 1$.*

**Proof of Proposition C.5:** Let $U$ denote the $t=0$ discounted value of future consumption. Lemma 1 implies that consumption at each time is proportional to output, which is in turn proportional to total flow profits earned by firms. Hence the discounted value of future consumption is proportional to the discounted value of future profits, or equivalently the discounted value of all firms:

$$\begin{aligned}
\frac{\bar{\pi}}{\bar{C}}U =& \int_{z_{A0}}^{\infty}\left[\frac{1}{\rho+\delta-\lambda\eta_A s}z_A\sigma_E K_A(0)+\frac{1}{\rho+\delta}\sigma_E K_B(0)\right]dH_A^*(z_{A0}) \\
&+\int_{1}^{z_{A0}}\left[\frac{1}{\rho+\delta}z_A\sigma_E K_A(0)+\frac{1}{\rho+\delta-\lambda\eta_B s}\sigma_E K_B(0)\right]dH_A^*(z_{A0}) \\
&+\delta N\int_{0}^{\infty}\exp(-\rho t)\left[\frac{1}{\rho+\delta}\sigma_E K_A(t)+\frac{1}{\rho+\delta-\lambda\eta_B s}\sigma_E K_B(t)\right]\mathbb{1}\left[\kappa(t)\geq\kappa^E\right]dt \\
&+\delta N\int_{0}^{\infty}\exp(-\rho t)\left[\frac{1}{\rho+\delta-\lambda\eta_A s}\sigma_E K_A(t)+\frac{1}{\rho+\delta}\sigma_E K_B(t)\right]\mathbb{1}\left[\kappa(t)<\kappa^E\right]dt.
\end{aligned}$$

This function is continuous in $z_{A0}$ except at the critical value that separates convergence to the BGP for technology $A$ from convergence to the BGP for technology $B$. To simplify, I assume that the social planner can choose the asymptotic direction of innovation at this value, so that the planner's problem (C8) is guaranteed to have a solution for $z_{A0}\in[1,\infty]$. Clearly higher values of $K_B(0)$ raise the value of innovating for $B$ relative to $A$, so the existence of the threshold $\hat{\kappa}$ is immediate given the linearity of the dynamical system (C2, C3) in $K$ and the linearity of the initial conditions (C4) in $K(0)$.

Suppose $\kappa(0)\geq\hat{\kappa}$, so that the social planner's solution $\hat{z}_{A0}$ is such that the economy converges to the BGP for $B$ as $t\uparrow\infty$. If $\kappa(0)<\kappa^*$, so that in equilibrium the economy converges back to the BGP for $A$, then we immediately have $\hat{z}_{A0}>z_{A0}^*$. Suppose then that $\kappa(0)\geq\hat{\kappa},\kappa^*$. We must have $\kappa(t)\geq\kappa^E$ for all $t\geq 0$ as entrants innovate for technology $B$, so given $z_{A0}$ social welfare at $t=0$ is

$$\begin{aligned}
\frac{\bar{\pi}}{\bar{C}}U =& \int_{z_{A0}}^{\infty}\left[\frac{1}{\rho+\delta-\lambda\eta_A s}z_A\sigma_E K_A(0)+\frac{1}{\rho+\delta}\sigma_E K_B(0)\right]dH_A^*(z_{A0}) \\
&+\int_{1}^{z_{A0}}\left[\frac{1}{\rho+\delta}z_A\sigma_E K_A(0)+\frac{1}{\rho+\delta-\lambda\eta_B s}\sigma_E K_B(0)\right]dH_A^*(z_{A0}) \\
&+\delta N\int_{0}^{\infty}\exp(-\rho t)\left[\frac{1}{\rho+\delta}\sigma_E K_A(t)+\frac{1}{\rho+\delta-\lambda\eta_B s}\sigma_E K_B(t)\right]dt.
\end{aligned}$$

Making use of Lemma C.3, we find

$$\int_0^\infty \exp(-\rho t) K_A(t)dt = K_A(0)\left(\frac{c_{A1}}{\rho} - \frac{c_{A2}}{\rho+\delta-\lambda\eta_A s}\right),$$
$$\int_0^\infty \exp(-\rho t) K_B(t)dt = K_B(0)\left(\frac{c_{B1}}{\rho+\delta+g_B^*-\lambda\eta_B s} + \frac{c_{B2}}{\rho-g_B^*}\right).$$

The integration constants $c_{A1}$, $c_{A2}$, $c_{B1}$, and $c_{B2}$ depend on $z_{A0}$ and are described in the proof of Lemma C.3. For our purposes, their derivatives satisfy

$$\frac{\partial c_{A1}}{\partial z_{A0}} = -\frac{\lambda\eta_A s\sigma_E}{\delta-\lambda\eta_A s} z_{A0} h_A^*(z_{A0}),$$
$$\frac{\partial c_{A2}}{\partial z_{A0}} = \frac{\partial c_{A1}}{\partial z_{A0}},$$
$$\frac{\partial c_{B1}}{\partial z_{A0}} = -\frac{\lambda\eta_B s\sigma_E}{2g_B^*+\delta-\lambda\eta_B s} h_A^*(z_{A0}),$$
$$\frac{\partial c_{B2}}{\partial z_{A0}} = -\frac{\partial c_{B1}}{\partial z_{A0}}.$$

Differentiating social welfare in $z_{A0}$, we have

$$\begin{aligned}\frac{\partial U}{\partial z_{A0}} \propto & \frac{1}{\rho+\delta} z_{A0}\sigma_E K_A(0) + \frac{1}{\rho+\delta-\lambda\eta_B s}\sigma_E K_B(0) - \left[\frac{1}{\rho+\delta-\lambda\eta_A s} z_{A0}\sigma_E K_A(0) + \frac{1}{\rho+\delta}\sigma_E K_B(0)\right] \\ & + \frac{\delta N\sigma_E K_A(0)}{\rho+\delta}\left(\frac{1}{\rho} - \frac{1}{\rho+\delta-\lambda\eta_A s}\right)\frac{1}{h_A^*(z_{A0})}\frac{\partial c_{A1}}{\partial z_{A0}} \\ & + \frac{\delta N\sigma_E K_B(0)}{\rho+\delta-\lambda\eta_B s}\left(\frac{1}{\rho-g_B^*} - \frac{1}{\rho+\delta+g_B^*-\lambda\eta_B s}\right)\frac{1}{h_A^*(z_{A0})}\frac{\partial c_{B2}}{\partial z_{A0}}.\end{aligned}$$

Note that the right side is strictly decreasing in $z_{A0}$, so that the first-order condition $\frac{\partial U}{\partial z_{A0}} = 0$ is both necesasry and sufficient to characterize the planner's solution $\hat{z}_{A0}$. Evaluating the expression above at $z_{A0}^*$, I observe that the first line collapses to zero by the definition of $z_{A0}^*$.

Hence

$$
\begin{aligned}
\left.\frac{\partial U}{\partial z_{A0}}\right|_{z^*_{A0}} &\propto \frac{1}{\rho+\delta}\left(\frac{1}{\rho}-\frac{1}{\rho+\delta-\lambda\eta_A s}\right)\frac{1}{h^*_A(z^*_{A0})}\left.\frac{\partial c_{A1}}{\partial z_{A0}}\right|_{z^*_{A0}} \\
&\quad+\frac{\kappa(0)}{\rho+\delta-\lambda\eta_B s}\left(\frac{1}{\rho-g^*_B}-\frac{1}{\rho+\delta+g^*_B-\lambda\eta_B s}\right)\frac{1}{h^*_A(z^*_{A0})}\left.\frac{\partial c_{B2}}{\partial z_{A0}}\right|_{z^*_{A0}} \\
&=\frac{1}{\rho+\delta}\left(\frac{1}{\rho+\delta-\lambda\eta_A s}-\frac{1}{\rho}\right)\frac{\lambda\eta_A s\sigma_E}{\delta-\lambda\eta_A s}z^*_{A0} \\
&\quad+\frac{\kappa(0)}{\rho+\delta-\lambda\eta_B s}\left(\frac{1}{\rho-g^*_B}-\frac{1}{\rho+\delta+g^*_B-\lambda\eta_B s}\right)\frac{\lambda\eta_B s\sigma_E}{2g^*_B+\delta-\lambda\eta_B s} \\
&\propto -\frac{1}{\rho+\delta}\frac{1}{\rho}\frac{\eta_A}{\rho+\delta-\lambda\eta_A s}z^*_{A0}+\frac{1}{\rho-g^*_B}\frac{1}{\rho+\delta+g^*_B-\lambda\eta_B s}\frac{\eta_B}{\rho+\delta-\lambda\eta_B s}\kappa(0).
\end{aligned}
$$

Using the definition $z^*_{A0}=\frac{\kappa(0)}{\kappa^E}$, we have

$$
\left.\frac{\partial U}{\partial z_{A0}}\right|_{z^*_{A0}} \propto \frac{1}{\rho-g^*_B}\frac{1}{\rho+\delta+g^*_B-\lambda\eta_B s}-\frac{1}{\rho+\delta}\frac{1}{\rho}.
$$

This value is strictly positive if and only if

$$
\rho(\rho+\delta)>\left(\rho-g^*_B\right)\left(\rho+\delta+g^*_B-\lambda\eta_B s\right).
$$

Simplifying yields

$$
g^*_B\left(g^*_B+\delta-\lambda\eta_B s\right)>-\rho\lambda\eta_B s.
$$

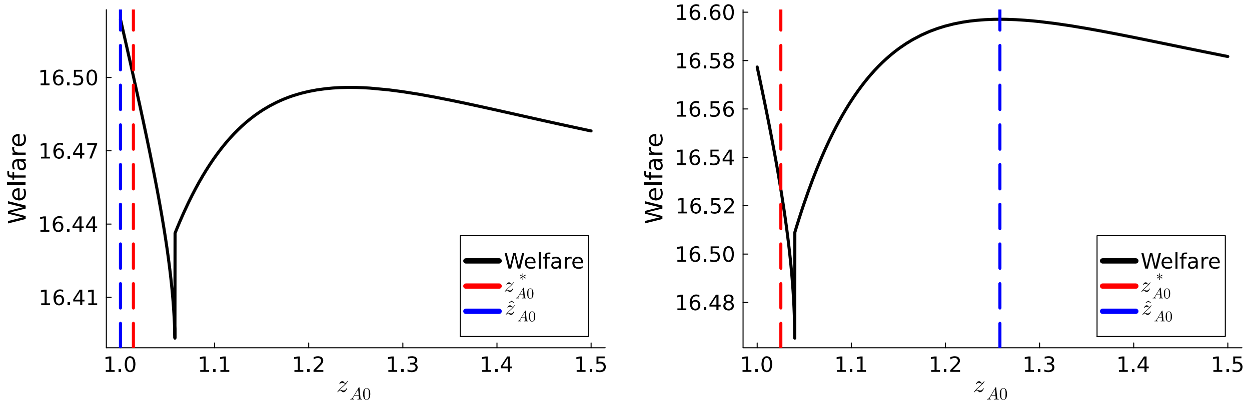
This inequality always holds because of the assumed lower bound (14) on the exit rate $\delta$. We can immediately conclude that $\hat{z}_{A0}>z^*_{A0}$ when $\kappa(0)\geq\hat{\kappa}$.

The remaining statement of the proposition follows by a symmetric argument: Whenever the social planner chooses $\hat{z}_{A0}$ so that the economy converges back to the BGP for technology $A$, knowledge spillovers lead the social planner to require more initial incumbents to innovate for technology $A$ than in equilibrium. ■

In general, the transition thresholds for the social planner $\hat{\kappa}$ and the equilibrium $\kappa^*$ cannot be ranked. This holds because the social planner internalizes knowledge spillovers on future entrants when choosing the long-run direction of innovation, but these spillovers are not necessarily always larger for a given technology: Technology $B$ spillovers are larger in the long-run given $\eta_B>\eta_A$, but technology $A$ spillovers may be larger in the short-run given incumbents' initial expertise for technology $A$ (i.e., the initial distribution $H^*_A$). However, Proposition C.5

Figure C.3: Social Welfare $U$ as a Function of the Cutoff $z_{A0}$

(a) Low $\kappa(0)$ (b) High $\kappa(0)$



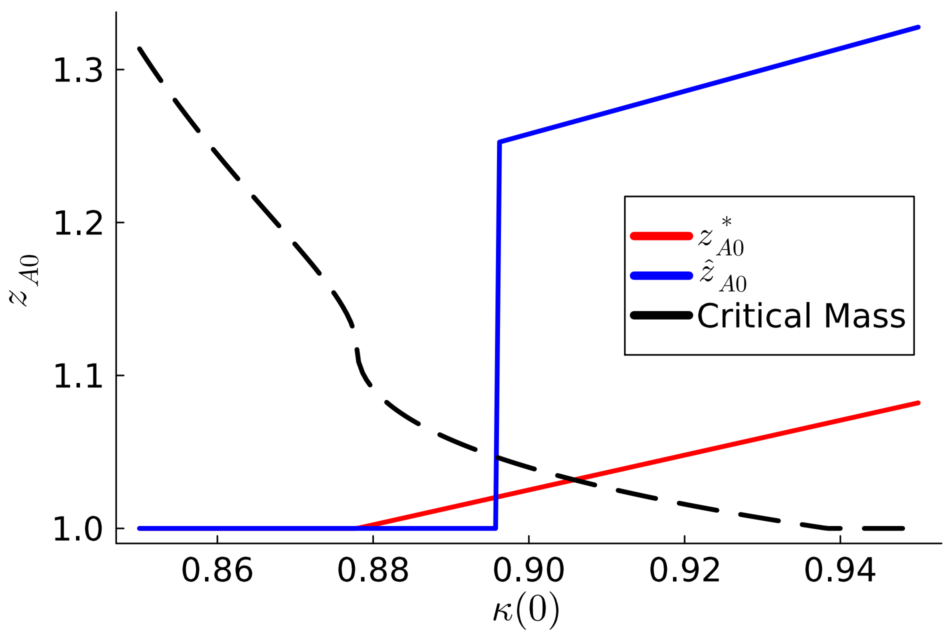

*Notes:* In each figure, I plot the social planner's objective (social welfare $U$) as a function of the initial cutoff $z_{A0}$. I also note the equilibrium cutoff $z^*_{A0}$ by a red dashed line and the efficient cutoff $\hat{z}_{A0}$ by a blue dashed line. The "Low $\kappa(0)$" figure sets $\kappa(0) = 0.89$, while the "High $\kappa(0)$" figure sets $\kappa(0) = 0.90$. All remaining parameters are exactly as in Figure C.1.

shows that for a given long-run innovation direction, the social planner always prefers to direct greater initial innovation in that direction than in equilibrium.

To provide a sense of the basic non-convexity in the social planner's problem (C8), Figure C.3 displays the objective $U$ as a function of the cutoff $z_{A0}$ for two different values of the initial knowledge stock ratio $\kappa(0)$. In both cases, social welfare attains a global minimum at the "critical mass" value of $z_{A0}$ that divides convergence to the BGP for technology $A$ from convergence to the BGP for technology $B$. On either side of that point, social welfare attains a global maximum, and the social planner's choice of the long-run innovation direction reduces to comparing the maxima from each side. In both cases, the equilibrium cutoff $z^*_{A0}$ leads to convergence back to the BGP for technology $A$. When $\kappa(0)$ is low, the social planner also chooses to innovate for technology $A$ in the long run, but consistent with Proposition C.5 the optimal cutoff $\hat{z}_{A0}$ is lower than that in equilibrium. When $\kappa(0)$ is high, the social planner instead chooses to innovate for technology $B$ in the long run.

Figure C.4 plots the equilibrium and efficient cutoffs $z^*_{A0}$ and $\hat{z}_{A0}$ for different values of $\kappa(0)$. In this example, the social planner prefers to transition to technology $B$ more often than in equilibrium ($\hat{\kappa} < \kappa^*$).

Figure C.4: Equilibrium and Efficient Cutoffs $z_{A0}$




*Notes:* The dashed black line denotes the cutoff $z_{A0}$ above which the economy converges to the BGP for technology $B$. The thresholds $\kappa^*$ and $\hat{\kappa}$ are the values of $\kappa(0)$ where $z^*_{A0}$ and $\hat{z}_{A0}$ intersect this line, respectively. The remaining parameters are exactly as in Figure C.1.

# D Data

This appendix describes the data for the analyses of Section 2 and Section 5.

## D.1 Data Sources

**Patents:** PatentsView, USPTO-granted patents 1980-2023

- Download date: 08 Oct 2024
- Exclusions:
  - non-utility, withdrawn, reissued patents
  - patents with missing assignee, CPC, filing year, or grant year data
  - patents assigned to Ethicon, Inc. (Johnson & Johnson subsidiary with anomalous forward citation counts, second only to IBM)
- Title and abstract text are processed using the NLTK package in Python to remove standard stop words, punctuation, numbers, and extra white space

**US Public Firm – Patent Match:** DISCERN 2.0 (Arora et al., 2024)

- Download date: 29 Sept 2024

**US Public Firm Financials:** Compustat North America Fundamentals Annual

- Download date: 20 Sept 2024

**New Technologies:** Kalyani et al. (2023)

- Download date: 14 Sept 2024

**Aggregate US R&D Expenditures:** "Research and Development: US Trends and International Comparisons," National Science Board (NSB-2024-6)

- Download date: 20 Sept 2024

**US GDP Deflator/GDP per Capita:** Federal Reserve Economic Data (FRED)

## D.2 Data Build: mRNA Case Study

**mRNA Therapy Patents.** I identify patents related to mRNA therapies by keyword search. I start with all patents that mention at least one term from each of the following lists:

mRNA terms: "mrna", " rna", "rna ", " ribonucleic"

therapy terms: "therap", "treat", "vaccin ", "innocul", "immun"

I then exclude patents that mention terms related to recombinant DNA/RNA or other types of RNA, which are involved in treatment technologies distinct from mRNA technology:

exclusion terms (RNA): "recombin", "rna interfer", "irna ", " rnai", "mirna", "sirna", "dsrna", "trna", "transfer rna", "double stranded rna", "small interfering rna", "double-stranded rna", "small-interfering rna", "micro rna", "micro-rna", "microrna", "reduce expression", "reducing expression", "inhibit expression", "inhibiting expression", "modulate expression", "modulating expression"

This procedure identifies 3408 mRNA therapy patents granted between 1980 and 2023 to 1211 unique assignees (as identified by PatentsView). Figure D.1(a) displays the number of patents filed and granted over 1980-2023.

**Conventional Vaccine Patents.** I identify patents related to conventional vaccines by keyword search. I start with all patents that mention at least one term from the following list:

vaccine terms: "vaccin", "innocul", "immuniz"

I then exclude patents that mention any of the exclusion terms for RNA technologies noted above. I also exclude any patents that mention terms related to cancer, because cancer vaccines constitute a distinct technology from conventional vaccines for infectious diseases:

exclusion terms (cancer): "cancer", "tumor", "tumour", "oncolog", "oncogen", "malign", "mestast", "neoplas"

Finally, I exclude remaining patents found in the set of mRNA therapy patents constructed above. This procedure identifies 9868 conventional vaccine patents granted between 1980 and 2023 to 2683 unique assignees (as identified by PatentsView). Figure D.1(b) displays the number of patents filed and granted over 1980-2023.

**Incumbent Patents.** I determine the set of mRNA therapy and conventional vaccine patents assigned to any of the following top 20 pharmaceutical firms by keyword search on patent assignees:

Top 20 Pharma: Johnson & Johnson, Sinopharm, Roche, Merck, Pfizer, AbbVie, Bayer, Sanofi, AstraZeneca, Novartis, Bristol-Myers Squibb, GSK, Eli Lilly Novo Nordisk, Shanghai Pharmaceuticals, Takeda, Amgen, Boehringer Ingelheim, Gilead Sciences, Siemens Healthineers

I search patent assignees for keywords related to each of these firms, including the names of any major subsidiaries or recently acquired firms. This procedure identifies 243 mRNA therapy patents and 1385 conventional vaccine patents granted to the firms above.

**Entrant Patents.** I determine the set of mRNA therapy and conventional vaccine patents assigned to Moderna, BioNTech, CureVac, or RNARx by keyword search on patent assignees. To more fully capture these firms' expertise, I also include all patents that list one of their founders as an inventor. This procedure identifies 242 mRNA therapy patents and 47 conventional vaccine patents granted to the four entrant firms or their founders.

**Forward Citations.** I compute the number of forward citations received by each patent from any other patent in the PatentsView dataset. These counts suffer from truncation bias that becomes severe for recently-granted patents. To control for this, I normalize each patent's forward citations by the average number of forward citations received by any USPTO patent granted in the same year. The normalized citation count of a patent $p$ granted in year $t$ is then

$$Count_p \equiv FCites_p \times \frac{|\{p' \text{ granted in } t\}|}{\sum_{p' \text{ granted in } t} FCites_{p'}}.$$

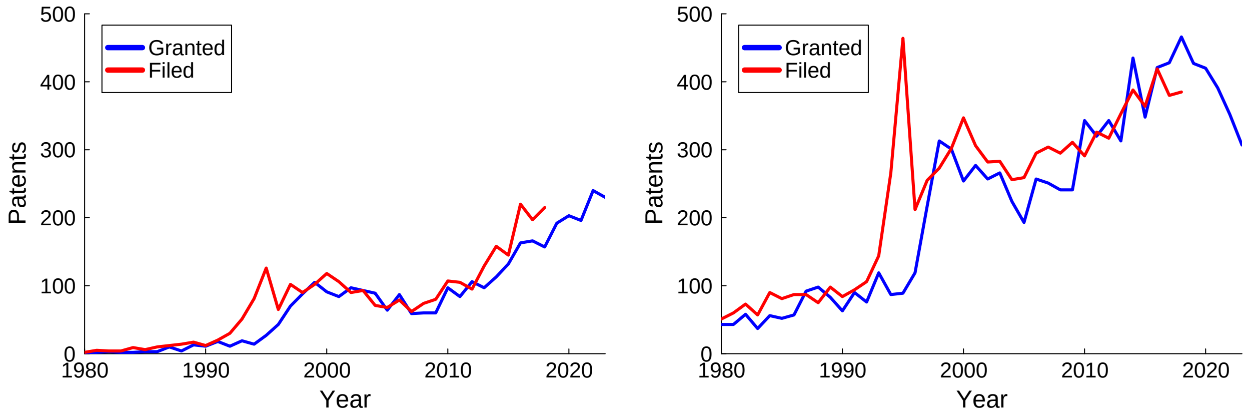
The average value of $Count_p$ across all mRNA therapy patents is 1.84, while the average across all conventional vaccine patents is 0.62.

**Sample.** On average, mRNA therapy and conventional vaccine patents each have a 3.5-year lag between filing and publication (see Figure D.2). Since I only obesrve granted patents, to avoid truncation bias I consider only patents filed through 2018 in Figure 2. Approximately

85% of each of mRNA therapy and conventional vaccine patents have a publication lag of five years or less.

Figure D.1: Patents Filed and Granted, 1980-2023

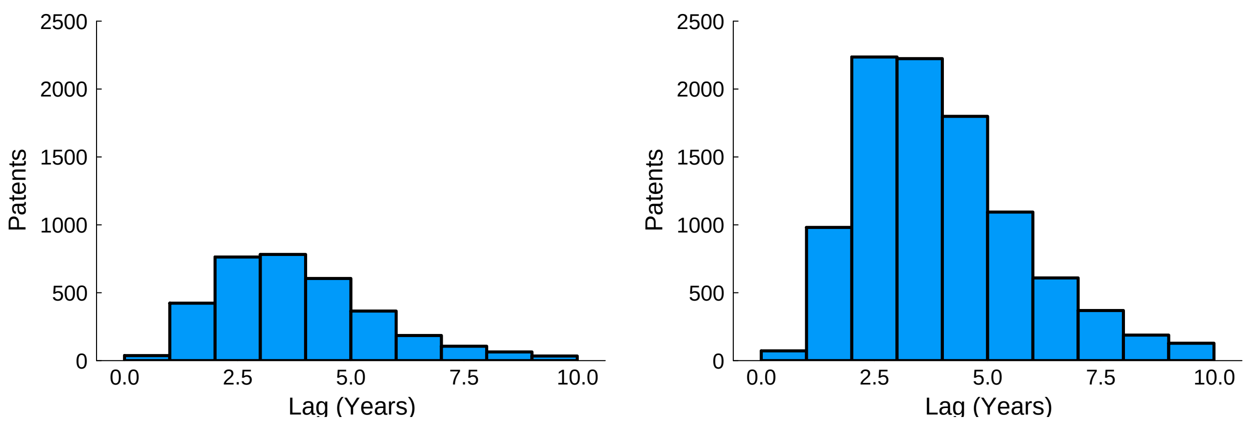
(a) mRNA Therapy Patents (b) Conventional Vaccine Patents




*Notes:* Given the lag between filing year and grant year documented in Figure D.2, I halt the “Filed” series in 2018. The spike in patent filings for conventional vaccines in 1995 coincides with implementation of the TRIPS agreement, which required WTO members to respect pharmaceutical patents.

Figure D.2: Lag from Patent Filing to Publication

(a) mRNA Therapy Patents (b) Conventional Vaccine Patents




*Notes:* For both sets of patents, over 85% of patents have a lag from filing to publication of five years or less.

## D.3 Data Build: Empirical Analysis

**Firm-Patent Match.** I use the DISCERN 2.0 dataset to recover all USPTO patents granted to US public firms from 1980 to 2021 (Arora et al., 2024). Each patent is matched to each of its original assignees; any firm with at least three consecutive years of coverage in Compustat North America is a potential match. I identify a firm with a unique *permno_adj* identifier defined in DISCERN 2.0, which is dynamically linked to *gvkey* identifiers in Compustat to allow consistent access to a firm's financial records over time (for example, accounting for name changes or acquisitions that could result in multiple *gvkeys*). The set of patents associated with a firm is the set of patents associated with the corresponding *permno_adj* identifier at the time of patent publication. I currently do not re-assign past patents associated with one *permno_adj* identifier to another after an acquisition, potentially biasing downward my estimates of firm knowledge stocks.

**Forward Citations.** I use yearly citation-weighted patent counts to measure a firm's innovative activity. The use of forward citations as a proxy for patent quality comes with two well-known limitations (e.g. Lerner and Seru, 2022): First, patents can only be cited after publication, so patents with more recent publication dates have mechanically fewer forward citations on average. Second, citation rates across different technology areas may be different independent of average patent quality (e.g., area-specific norms, differences in the combinatorial nature of innovation, etc.). As a result, raw citation counts may not yield a measure of quality comparable across technology areas.

I address these concerns with the following two-step normalization procedure. First, for each year $t$, I compute the mean number of forward citations $FCites_p$ across all patents $p$ granted in year $t$:

$$\overline{FCites}_t \equiv \frac{\sum_{p \text{ granted in } t} FCites_p}{|\{p \text{ granted in } t\}|}.$$

This mean is plotted for each year $t$ from 1980 to 2023 in Figure D.3(a). I define $FCites_{p1}$ for each patent $p$ granted in year $t$ as the normalized forward citation count

$$FCites_{p1} \equiv \frac{FCites_p}{\overline{FCites}_t}.$$

This step provides a simple correction for the truncation issue noted above. But I note that it can also introduce a bias of its own if citation counts are particularly high in a given year because a large number of high-quality patents were granted: This normalization effectively equalizes the total measure of quality across all grant years.

Second, for each CPC subclass $s$, I compute the mean of $FCites_{p1}$ across all patents $p$ assigned to class $s$:

$$\overline{FCites}_{s1} \equiv \frac{\sum_{p \in s} FCites_{p1}}{|\{p \in s\}|}.$$

The distribution of this mean across subclasses $s$ is plotted in Figure D.3(b). I define $FCites_{p2}$ for each patent $p$ as the normalized value of $\overline{FCites}_{p1}$, where the normalizing factor is the mean of $\overline{FCites}_{s1}$ across all subclasses $s$ to which $p$ is assigned:

$$FCites_{p2} \equiv FCites_{p1} \frac{|\{s \ni p\}|}{\sum_{s \ni p} \overline{FCites}_{s1}}.$$

This step corrects for differences in citation rates across CPC subclasses that are uniform over time. It can introduce bias if citation counts are particularly high in a given CPC subclass because a large number of high-quality patents were assigned to that subclass over the sample: This normalization effectively equalizes the total measure of quality across CPC subclasses over the sample.

Throughout the empirical analysis of Section 5, the "citation weight" of a patent is equal to $FCites_{p2}$. This normalization procedure is weaker than that used by, for example, Hall et al. (2001) and Kalyani et al. (2023), which normalizes citations within each class-year. My normalization allows the total measure of quality to vary across subclasses at each time and over time with a subclass, ensuring that it can reflect "innovation bursts" for particular subclasses over a period of several years (e.g., for personal computing-related classes in the 1980s and 1990s). It is the minimal normalization that plausibly allows comparability between patents granted at different times and in different technology areas.

**Citation-Weighted Patents.** The citation-weighted patent flow for firm $i$ in year $t$ is simply the sum of $FCites_{p2}$ across all patents filed in year $t$ and assigned to firm $i$:

$$Pat_{it} \equiv \sum_{\{p \text{ filed in } t \text{ by } i\}} FCites_{p2}.$$

Figure D.4 displays raw and citation-weighted patent counts by year, both for the mean across firms in the sample of Section 5 and for a particular firm (Microsoft).

**Knowledge Stocks.** As noted in Section 5, firm $i$'s generic internal knowledge stock at time $t$ is constructed by the perpetual inventory method from the patent flows $Pat_{it}$:

$$K_{it}^{\text{Firm}} \equiv (1-\nu)K_{it-1}^{\text{Firm}} + Pat_{it-1}.$$

If firm $i$ becomes public in year $t_0$, I initialize by setting $K_{it_0} = 0$ because I do not observe the firm's previous patenting behavior. To help address this measurement error, each regression in Section 5 is restricted to firms that have been public for at least one year and includes fixed effects for the number of years for which the firm has been publicly listed.

To construct the aggregate knowledge stock $K_{it}^{\text{Agg}}$, I first compute the share of each firm $i$'s patents assigned to each CPC subclass $s$ over the sample:[41]

$$\omega_{is} \equiv \frac{|\{p \in i \cup s\}|}{|\{p \in i\}|}.$$

Since a single patent is generally assigned to many subclasses, the weights $\omega_{is}$ do not sum to one across $s$. I defined normalized weights

$$\tilde{\omega}_{is} \equiv \frac{\omega_{is}}{\sum_{s'} \omega_{is'}}.$$

I also compute the aggregate generic knowledge stock for each CPC subclass $s$ from the yearly flows of patents $Pat_{st}$ filed in subclass $s$:

$$K_{st}^{\text{Agg}} \equiv K_{st-1}^{\text{Agg}} + Pat_{st-1},$$

with initial value $K_{s1980} = 0$. I then construct firm $i$'s aggregate knowledge stock as the patent share-weighted average

$$K_{it}^{\text{Agg}} \equiv \sum_{s} \tilde{\omega}_{is} K_{st}^{\text{Agg}}.$$

Firm $i$'s internal knowledge stock for technology $\theta$ at time $t$ is defined analogously to the generic knowledge stock $K_{it}^{\text{Firm}}$:

$$K_{i\theta t}^{\text{Firm}} \equiv (1-\nu)K_{i\theta t-1}^{\text{Firm}} + Pat_{i\theta t-1},$$

[41] I perform this step of the construction at the subclass level to ensure a more precise characterization of the knowledge relevant for each firm's innovation.

where

$$P_{i\theta t} \equiv \sum_{\{p \text{ filed in } t \text{ for } \theta \text{ by } i\}} FCites_{p2}.$$

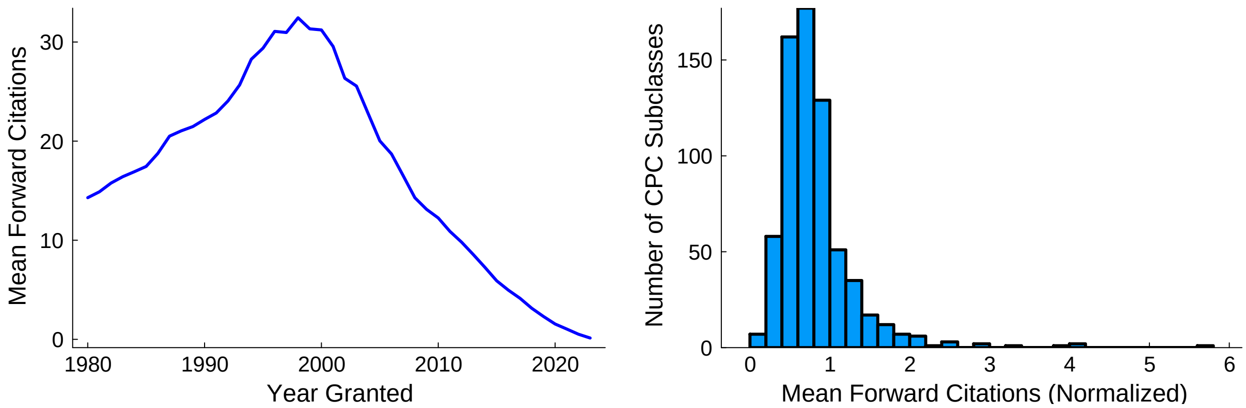
If firm $i$ becomes public in year $t_0$, I initialize by setting $K_{i\theta t_0} = 0$.

The aggregate knowledge stock for technology $\theta$ at time $t$ is simply constructed by the perpetual inventory method from the yearly flows of patents $Pat_{\theta t}$ filed for technology $\theta$:

$$K^{\text{Agg}}_{\theta t} \equiv (1-\nu)K^{\text{Agg}}_{\theta t-1} + Pat_{\theta t-1}.$$

Figure D.3: Citation Normalizing Factors

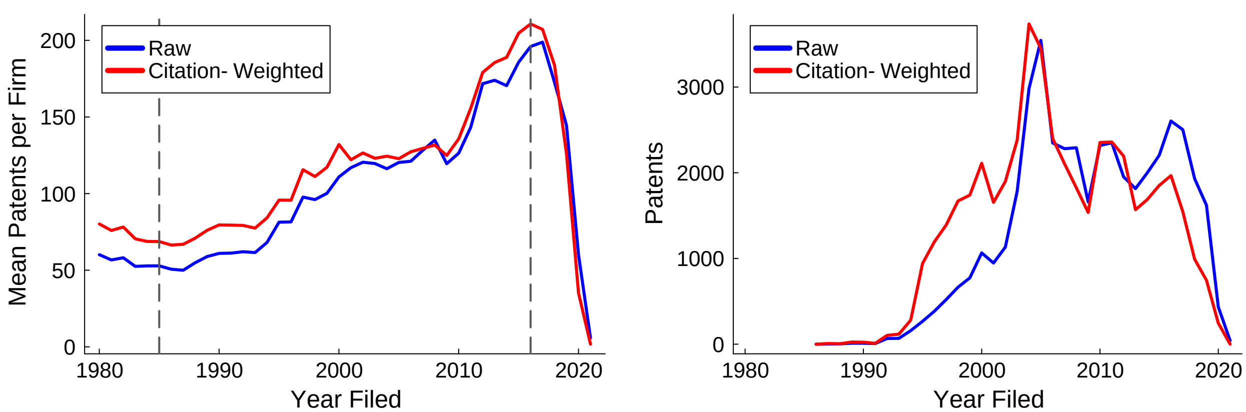
(a) $\overline{FCites}_t$ (b) $\overline{FCites}_{s1}$




*Notes:* Figure D.3(a) displays the mean number of forward citations across all patents granted in each year $t$, $\overline{FCites}_t$. Figure D.3(b) displays the distribution (across subclasses) of the mean number of (year-normalized) forward citations for all patents in a subclass, $\overline{FCites}_{s1}$.

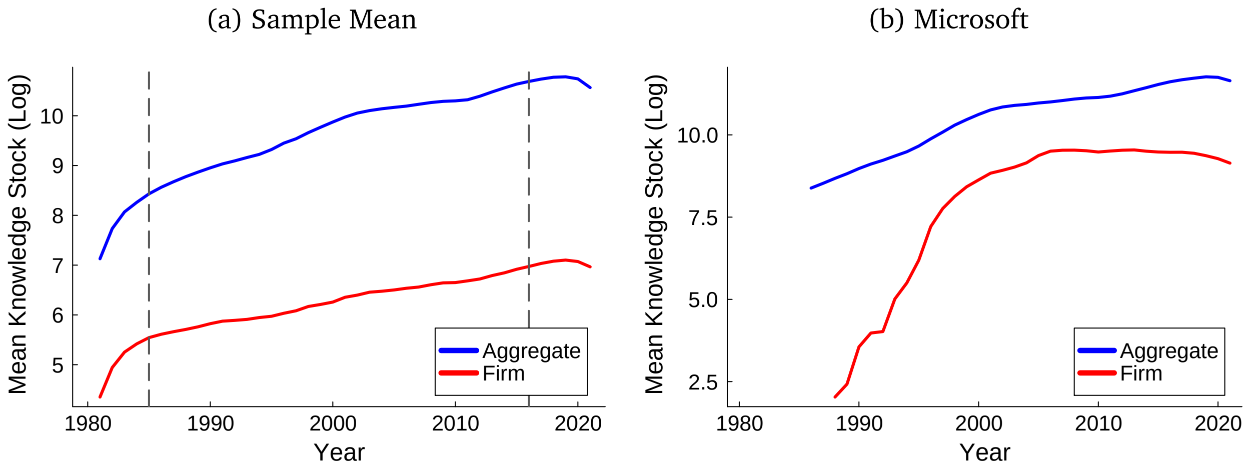
Figure D.4: Patent Counts, 1980-2021

(a) Sample Mean (b) Microsoft




*Notes:* Figure D.4(a) displays the mean number of patents (raw and citation-weighted) filed each year across firms in the sample used in Section 5. Figure D.4(b) displays the corresponding patent counts for Microsoft. The vertical dashed lines in Figure D.4(a) denote the beginning (1985) and end (2016) of the sample period for the regressions in Section 5. Microsoft becomes publicly listed in 1986.

Figure D.5: Generic Knowledge Stocks, 1980-2021




*Notes:* Figure D.5(a) displays the log of the mean firm and aggregate knowledge stocks each year across firms in the sample used in Section 5. Figure D.5(b) displays the corresponding log knowledge stocks for Microsoft. The vertical dashed lines in Figure D.5(a) denote the beginning (1985) and end (2016) of the sample period for the regressions in Section 5. Microsoft becomes publicly listed in 1986.

# E Empirics: Additional Results and Robustness

Table E.1: Regression Results: Technology Patenting after Emergence (34)

| | Technology Patents | | Technology Patent Share | |
|---|---|---|---|---|
| | (1) Full Sample | (2) No Early Patents | (3) Full Sample | (4) No Early Patents |
| $\log(K^{\text{Firm}}_{iT_\theta})$ | -0.0828*** | -0.0796*** | -0.0024*** | -0.0014** |
| | (0.0315) | (0.0267) | (0.0009) | (0.0007) |
| $\log(K^{\text{Agg}}_{iT_\theta})$ | -0.5266*** | -0.5038*** | -0.0060 | -0.0059 |
| | (0.1738) | (0.1609) | (0.0050) | (0.0060) |
| $\log(s_{iT_\theta})$ | 0.0041 | -0.0054 | -0.0004 | -0.0005 |
| | (0.0116) | (0.0115) | (0.0005) | (0.0004) |
| $\log(K^{\text{Firm}}_{i\theta T_\theta})$ | 0.3175*** | 0.3784*** | 0.0061*** | 0.0082* |
| | (0.0232) | (0.1143) | (0.0009) | (0.0049) |
| $\log(K^{\text{Agg}}_{\theta T_\theta})$ | 0.1951*** | 0.1847*** | 0.0033*** | 0.0029*** |
| | (0.0129) | (0.0141) | (0.0004) | (0.0005) |
| $R^2$ | 0.63610 | 0.58407 | 0.28643 | 0.25438 |
| Observations | 13,662 | 11,002 | 13,662 | 11,002 |
| Dep. Var. Mean | 2.3309 | 2.0265 | 0.01771 | 0.01582 |
| Dep. Var. SD | 2.2448 | 2.1524 | 0.04604 | 0.04441 |

*Significance: *** 0.01, ** 0.05, * 0.1*

*Notes:* All regressions include fixed effects by firm, emergence year, and the number of years for which the firm has been publicly listed. They also include dummy variables for zero values of each of the knowledge stocks and R&D expenditures $s_{it}$. All standard errors are clustered at the firm level.