Firms as Foragers*

Vasco M. Carvalho   Lukas B. Freund

July 11, 2026

Please click here for the latest version.

Abstract

We develop a theory of growth in which firms forage in idea space. A firm exploits a patch of related ideas, gradually exhausting opportunities for quality improvement, and then searches for a new patch. We cast this explore-exploit tradeoff as a tractable optimal-stopping problem and embed it in an endogenous-growth model. The composition of innovation—improving existing ideas versus discovering new ground—emerges as an equilibrium object. To construct an empirical representation of the idea space, we apply natural language processing to patent text data. The data support the theory's central premises: returns to local exploitation diminish; firms stay longer on richer patches; and entry into new patches yields more and better patents. We calibrate the model to U.S. data and establish two results, on the composition of growth and on its pace. First, at a twenty-year horizon, patenting in new clusters accounts for over half of growth from quality improvements: sustained growth rests on firms continually entering new territory. Second, the model sign-identifies the origins of the productivity slowdown of the last four decades: exploitation spells have not shortened, weighing against worsening exploitation and tentatively pointing to harder exploration.

*Carvalho: University of Cambridge & CEPR, vmcarvalho.web@gmail.com, www.vasco-m-carvalho.github.io; Freund: Boston College, lukas.beat.freund@gmail.com, www.lukasfreund.com. We thank Susanto Basu and Bill Janeway for helpful conversations. We gratefully acknowledge financial support from the Keynes Fund (project code JHXI). Carvalho gratefully acknowledges the financial support of ERC Micro2Macro No. 101001221. We thank Tianxiang Zheng and Claude for excellent research assistance. First version: July 10, 2026.

1 Introduction

Much like animals foraging for sustenance, firms face a continual choice between locally exploiting their current stock of knowledge and exploring for new ideas and revenue streams. This cycle of exploitation and exploration—drawing down a patch until it thins, then wagering on a new one—has been a central theme in the study of organizations (e.g., March, 1991).

Consider the storied account of Apple's reinvention in the 21st century. Its breakthroughs came as it entered new product categories: the iPod in 2001, the iPhone in 2007, and wearables such as the AirPods thereafter.[1] Between the jumps, each product was refined generation by generation. This history is legible in Apple's patents. Figure 1 follows three clusters of related inventions,[2] one behind each category: multimedia encoding behind the iPod, touch interfaces behind the iPhone, in-ear audio behind the AirPods. Panel (a) plots average forward citations to Apple's patents in each cluster. The three peaks arrive in sequence: media around the iPod, touch around the iPhone, earphones later still.[3] Revenues follow with a lag (panel (b)): the iPod rises through the 2000s and recedes; the iPhone holds near its peak far longer, a persistent plateau rather than a passing hump; and wearables climb as the iPhone matures. Each new category adds a stream of revenue that sustains the firm even as older ones decline.

These clusters—patches, in the forager's metaphor—are regions of a larger space of ideas, in which we position every patent by its text: similar ideas sit close together, distant ones far apart. Panel (c) traces Apple's annual position in this space: large strides through the 2000s, into digital media and touch interfaces, after which movement turns local, circling and refining rather than striking out. Panel (d) confirms the shift with a complementary gauge: the distance from each Apple patent to its nearest predecessor stays high through the early years and falls steadily from the mid-2000s. Read through its patents, Apple's history is a sequence of forays into new territory followed by spells of incremental improvement.

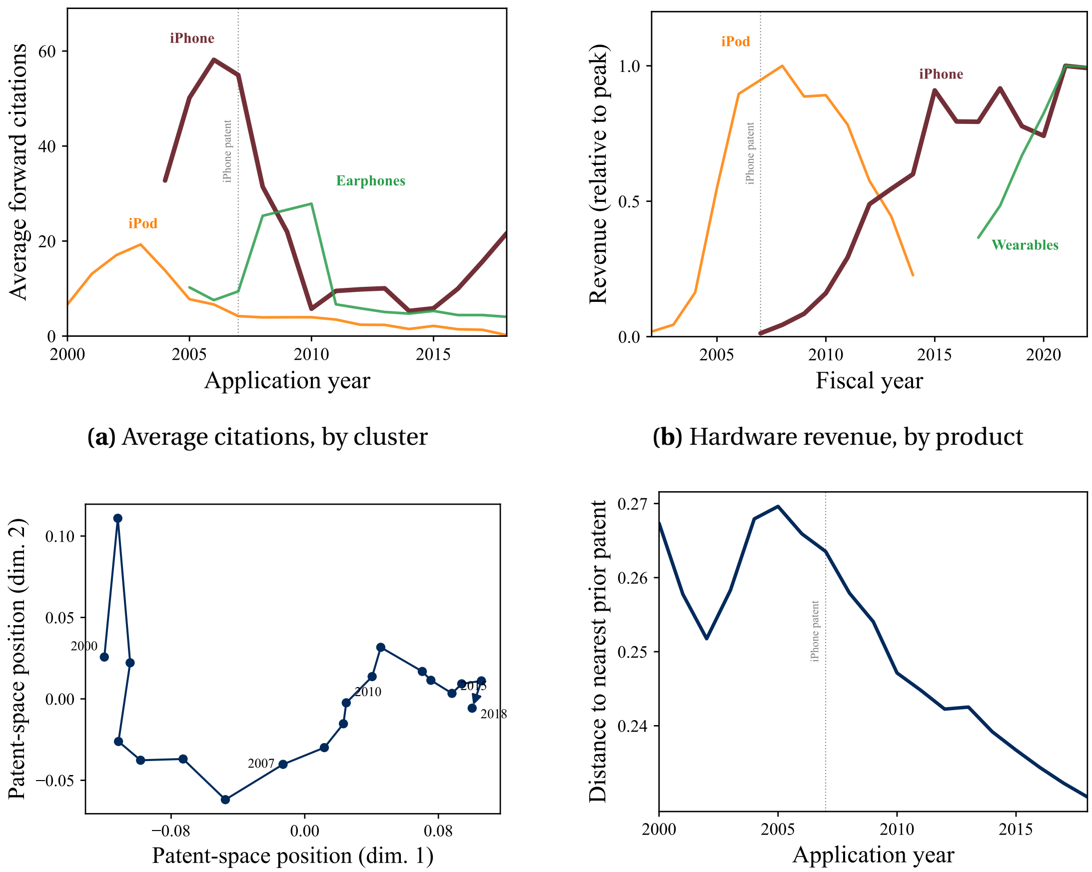
This cycle has no counterpart in the theory of growth: in workhorse models, the ideas behind a product line never run dry. The question of when a firm should move on—and with it, how this decision shapes the composition of innovation and the pace of aggregate growth—goes unasked.

In this paper we take up this question and develop an empirically grounded model in which firms forage in an idea space, placing the explore–exploit cycle at the center of a theory of economic growth. Using text embeddings of patents, we construct an empirical representation of idea patches at scale and establish three facts: returns to exploitation diminish within a patch;

[1] Steve Jobs himself held that one should not linger too long on a given product or area: "If you do something and it turns out pretty good, then you should go do something else wonderful, not dwell on it for too long. Just figure out what's next." (Interview with Brian Williams, NBC News, 25 May 2006).

[2] We embed each patent's title and abstract and cluster similar patents; Section 2 provides details.

[3] The most cited is U.S. Patent 7,479,949, the sweeping multitouch patent known in the press as "the Steve Jobs patent" after its first-named inventor.



(a) Average citations, by cluster

(b) Hardware revenue, by product

(c) Path through patent space

(d) Distance to nearest prior patent

Figure 1: Apple's innovation history as a sequence of exploration and exploitation

Notes. Apple patents are identified by CRSP permno 14,593. The figure tracks three emblematic embedding-based technology clusters: *iPod* (media management), *iPhone* (touch interface; including the famous patents 8,564,544 and 7,479,949), and *Earphones* (in-ear audio / AirPods). Panels (a), (c) and (d) use patent application years 2000–2018; panel (b) uses fiscal years 2002–2022. Panel (a) plots average five-year forward citations for each cluster year. Panel (b) plots Apple hardware net sales in real 2023 dollars, from 10-K filings, indexed to each product's own peak; the reported "Wearables" category includes AirPods. Panel (c) plots the yearly centroid of Apple's patent embeddings projected to two dimensions by principal components. Panel (d) plots the average distance to the nearest prior patent in a five-year window. Patent series are smoothed with a three-year moving average.

firms stay longer on richer patches; and entering new patches yields more and better innovation. Guided by these facts, and integrating ideas from optimal foraging theory in behavioral ecology, we build the model as follows. A firm improves an active product line by drawing on a patch of related ideas whose returns fall as the patch is depleted; the firm's central decision is when to abandon that patch and explore for a new one. We cast this decision as an optimal stopping problem, tractable enough to embed in an otherwise standard general-equilibrium model of endogenous growth. We characterize the firm's stopping rule analytically and decompose growth into within-patch research productivity and the share of time firms spend exploiting rather than exploring. Finally, we calibrate the model to address two questions. First, how much of growth comes from exploring new patches rather than exploiting old ones? Both model and data attribute half or more of twenty-year growth to forays into new territory. Second, does the U.S. productivity growth slowdown over the last few decades reflect worsening exploitation or harder exploration? The model yields a sign-identification strategy based on the duration of exploitation spells; taken to the data, it points away from worsening exploitation and tentatively toward harder exploration.

Empirical facts. Our first contribution is empirical. We combine four decades of U.S. patent records with firm-level financials. Using natural-language-processing tools to represent each patent's text as a vector embedding, we place every patent in a high-dimensional idea space and cluster nearby patents into discrete idea patches—the empirical counterpart to the regions of knowledge that firms exploit and move between. A firm's position in a given year is the location of its patents; the quality of its innovations is measured by patent-level proxies, including forward citations, private market value (Kogan *et al.*, 2017), breakthrough importance (Kelly *et al.*, 2021), and creativity (Kalyani, 2024). This yields, for each firm, a trajectory through the idea space.

We establish three facts. The first concerns what happens within a patch. We find that there are diminishing returns to exploitation: the quality of successive patents declines on every measure as a firm continues to innovate within a single region of the idea space, and a firm's best work tends to come early, with roughly a fifth of its most-cited patents arriving in the opening tenth of a spell. This decline is not driven by a fall in research input. The second fact concerns movement across patches. Firms facing these diminishing returns routinely enter new areas, and how long they stay reflects what a patch offers: spells are longer where returns are richer, with durations ranging from a median of one active year to runs lasting decades. The third fact concerns the resulting firm-level trajectories. A firm's path through the idea space is uneven, combining long stretches of small, local steps with occasional jumps into new territory; Apple's trajectory is the rule rather than the exception. These jumps pay off: entering a new patch raises the quantity and quality of a firm's patents and, with a lag, its sales.

Taken together, these facts describe a firm that faces diminishing returns within any patch, moves on when the patch is drawn down, and is rewarded for doing so. This is the explore–exploit tradeoff in the data, and it disciplines the theory we develop.

Theoretical framework. Our second contribution is theoretical. We integrate the central insight of optimal foraging theory—an agent exploiting a depleting resource must decide when to leave it for a richer one—into endogenous growth theory.

The model places this decision inside a continuous-time economy with expanding, quality-differentiated varieties. A firm develops a single active product line while collecting profits from a legacy portfolio of past products it no longer improves. To improve the active line, the firm draws on a patch of related ideas, characterized by its density: how much room for quality improvements the patch still holds. (At some point, the smartphone is about as good as a smartphone will get.) The firm exploits the patch through R&D, and improvements arrive stochastically, raising product quality. Two assumptions are central. First, improvements are larger the denser the patch. Second, exploiting a patch depletes it: density drifts down, so successive improvements grow smaller. The firm's problem is to choose when to abandon a depleting patch and explore for a new one whose density it cannot know in advance—an optimal stopping problem. On abandoning, the firm searches the idea space until it encounters a new patch and begins a fresh cycle. Aggregate growth is the accumulation of these cycles.

Our theory delivers two results. The first characterizes the firm's optimal policy: a single threshold in patch density, such that the firm exploits while the patch stays rich enough and abandons it once density falls to a cutoff. Under an approximation, we solve for this cutoff in closed form: the firm leaves precisely when the flow return to continued exploitation falls to the opportunity cost of staying—the return it could earn searching for a new patch, net of research costs. This is the economic counterpart of the marginal value theorem of optimal foraging theory (Charnov, 1976). We then show how firm-level quality growth aggregates to economy-wide growth, under conditions we make precise.

The second result concerns the pace of growth. Aggregate growth decomposes into two terms: how productively firms improve quality while exploiting a patch, and the share of time spent exploiting rather than exploring. Growth is slow for either of two reasons: little is gained from each patch, or too much time is lost exploring between them. We analytically characterize how both terms respond to the model's parameters. When patches deplete more quickly, firms abandon them at the same threshold but exhaust them sooner; they spend more time exploring, and growth falls because less of the economy's effort goes to exploitation. When new patches are easier to find, exploration becomes more valuable and firms abandon patches earlier, while still relatively rich; spells shorten, but growth typically rises, because exploitation is confined to the most productive part of each patch. When improvements arrive more often or are larger, exploitation becomes more productive and firms remain on their patches longer, so growth rises.

Applications. Our third contribution is a quantitative illustration of the model. We calibrate it to U.S. patent and firm data through a mix of a priori information, direct mapping to data moments, and simulated method of moments. We use the calibrated model to study two applied questions, one about the composition of growth and one about its pace.

Our first exercise concerns the composition of growth: how much of aggregate U.S. growth comes from exploration rather than exploitation? We make this precise as the share of quality-improvement growth originating in patches firms have newly entered. We find that, over a twenty-year horizon, newly entered patches account for roughly two-thirds of quality-improvement growth in the model and about half in the data. Continued exploration is thus a first-order source of growth, on a par with the exploitation of patches firms already hold. More generally, the framework shows that aggregate growth does not determine its composition. We construct counterfactual economies that grow at the same rate but draw very different shares of that growth from new patches—one rotating quickly through many patches, the other exploiting richer patches for longer.

Our second exercise offers a diagnostic for the origins of the U.S. productivity growth slowdown of the last four decades. In our model, a slowdown can originate in worsening exploitation or worsening exploration, and the growth rate alone cannot reveal which, since both lower it. The model, however, sign-identifies the two: worsening exploitation shortens the time firms spend on a patch and speeds their entry into new ones, whereas worsening exploration lengthens that time and slows entry. In U.S. data, average patch duration has, if anything, risen, and entry has fallen—a pattern that weighs against a slowdown driven by deteriorating exploitation and is tentatively consistent with harder exploration.

Finally, we discuss how the same diagnostic speaks to artificial intelligence as a method of invention. If AI mainly accelerates exploitation of technologies a firm already holds, firms stay on each patch longer—growth improves, but innovation narrows further to familiar territory. If instead AI makes new patches easier to find, firms enter new patches faster. Both cases raise the growth rate, so pace alone cannot reveal what kind of method of invention AI is. Its imprint on the time firms spend on a patch can.

Outline. The rest of the paper proceeds as follows. We close the introduction with a discussion of the related literature. Section 2 describes the data, our measurement of idea patches, and the three empirical facts. Section 3 develops the theory: the foraging environment, the balanced-growth equilibrium, and the closed-form characterization of the firm's stopping problem and growth. Section 4 calibrates the model and turns to the two applications. Section 5 concludes.

Related literature. Our paper contributes to three literatures: growth theory, empirical work on innovation and growth, and the study of the explore–exploit tradeoff across disciplines.

First, our theory relates to modern endogenous growth models in which growth emerges from firms' innovation decisions. Our theory features expanding, quality-differentiated varieties, using a firm structure as in Klette and Kortum (2004), with growth arising from quality improvements (Grossman and Helpman, 1991; Aghion and Howitt, 1992).[4] Like Garcia-Macia *et al.* (2019),

[4]See, among others, Acemoglu *et al.* (2018), Aghion *et al.* (2023, 2025), De Ridder (2024), Peters and Walsh (2026), Berlingieri *et al.* (2025).

we decompose growth into sources—but where they distinguish own-product improvement, creative destruction, and new varieties, inferred indirectly from employment dynamics, we distinguish by the vintage of patches, measured directly from patents. We also share conceptual ground with papers that model innovation as drawing from, moving through, or diffusing over a space of ideas.[5] In those models agents draw from or copy an existing distribution; our firms deplete local regions and choose when to relocate.[6]

Most closely related are recent papers that distinguish between types of innovation, notably Akcigit and Kerr (2018), Acemoglu *et al.* (2022), and Aghion *et al.* (2026). In Akcigit and Kerr (2018), multi-product firms choose internal R&D (improving own lines) and external R&D (acquiring new lines via creative destruction) as flow intensities under convex costs, so the equilibrium mix follows from static first-order conditions re-optimized each instant. Our theory differs in two respects: a state variable tracks how close the current patch is to exhaustion—in Akcigit and Kerr (2018), internal opportunities never deplete—and finding a new patch costs time in addition to R&D resources. Both features are essential to our applications. Empirically, depletion is a salient feature of the data, and in the model the speed of depletion relative to the time cost of exploration shapes the vintage composition of growth. Analytically, growth factors into within-patch productivity times an exploitation time share, so spell durations emerge as equilibrium objects with identifying content. We exploit this structure to diagnose the drivers of growth slowdowns and accelerations.

The notion of diminishing returns at the patch level also appears in Acemoglu *et al.* (2022), but there it motivates cross-sectional evidence on managers. In their model, redirecting innovation toward a new technology involves no costly search time, and the model is silent on the objects at the center of our analysis: how long firms exploit a patch, how they traverse the idea space, and the composition of growth. Finally, we share with Aghion *et al.* (2026, ABBB henceforth) the thesis that ideas get harder to find within any single technology while growth is sustained by turnover. Their mechanism, however, is deliberately stylized in two ways. Technology-specific exhaustion is a Poisson shock that switches a firm's development capacity from a constant drift to zero, and a strict division of labor—entrants do all research, incumbents only develop—means the clock resets exclusively through firm death and replacement. In our model, depletion is gradual and incumbents themselves switch between exploitation and exploration. Both features of our model are needed to speak to the vintage composition of growth as an endogenous object and to relate the model to data on innovation activity. Importantly, ABBB's conclusion explicitly

[5] An early precedent is Jovanovic and Rob (1990), along with Kortum (1997) and Weitzman (1998). More recent papers include Lucas and Moll (2014), Perla and Tonetti (2014), and Buera and Oberfield (2020).

[6] A complementary literature endogenizes which questions researchers pursue (Hopenhayn and Squintani (2021); Akcigit *et al.* (2021); Carnehl and Schneider (2025)). These papers develop the micro theory of individual researchers; our focus is firm dynamics and aggregate implications.

calls for confronting the mechanism with granular microdata on how the innovation clock resets. Our paper answers that call, including through evidence on diminishing returns at the patch level.

Second, we relate to empirical work on innovation and growth, specifically studies that measure innovation with patent data and document trends in research productivity (Bloom *et al.*, 2020; Jones, 2009; Park *et al.*, 2023; Fort *et al.*, 2026). Our contribution is twofold. We apply natural language processing to patent text to construct an empirical representation of the idea space; combined with patent-quality measures from Kogan *et al.* (2017, KPSS henceforth), Kelly *et al.* (2021), and Kalyani (2024), this yields new facts about firm innovation dynamics. Most importantly, we show that "ideas getting harder to find" operates at the level of the idea patch,[7] and that exploration acts as a force of renewal: entering new patches raises the quantity and quality of a firm's innovation (see also Carvalho *et al.* (2021)). Beyond documenting facts, we use the embeddings to construct the state space of a structural model—discrete patches with observed entry, spells, and exit—which the model imbues with identifying content. Theory and evidence together discriminate among explanations for the decline in research productivity and, prospectively, identify what kind of method of invention AI is. In contemporaneous work, Ganguli *et al.* (2026) also represent patents in an embedding space, but to a different end. They study a static model of spatial competition in which inventors spread out to escape crowding, and the resulting attenuation of spillovers lowers research productivity. We instead focus on when a firm abandons a depleting patch to search for a new one.

Third, the explore–exploit tradeoff at the heart of our paper runs through several fields: organizations and management, from Cyert and March (1963) and March (1991) to Christensen (1997), and formal single-agent models of search and experimentation (e.g., Weitzman, 1979; Callander, 2011). Our study draws especially on optimal foraging theory in behavioral ecology,[8] which we integrate into endogenous growth theory; the intermittent search structure follows Bénichou *et al.* (2005); Bénichou (2006); Bénichou *et al.* (2011). By casting the patch-leaving decision as a canonical optimal stopping problem, we retain enough tractability to embed it in market equilibrium with growth. Our threshold condition generalizes the marginal value theorem of Charnov (1976)—leave a patch when the marginal return falls to the average return of the environment—but our outside option is an equilibrium object rather than an exogenous intake rate. Finally, we take the foraging idea to firm-level innovation data at scale.

[7] Our evidence is complementary to Argente *et al.* (2024), who document that product sales decline steadily over the life cycle in consumer-goods scanner data, driven by obsolescence.

[8] See Stephens and Krebs (1986) for a comprehensive treatment and Charnov (1976) on the marginal value theorem specifically. Classic empirical tests of the marginal value theorem include Krebs *et al.*'s (1974) study of captive black-capped chickadees foraging in a large aviary for small pieces of mealworm hidden in artificial pine cones and Cowie's (1977) study of how great tits' patch residence time responds to travel time.

2 Empirical motivation

We establish three motivating facts about the dynamics of firm innovation: There are diminishing returns to exploitation at the level of "patches of ideas"; firms move around such patches, which vary in quality; and this behavior yields a dynamic sequence of firm-level spatial movements and rewards. This section first describes our empirical methodology (Section 2.1), then describes each fact in detail (Section 2.2).

2.1 Methodology

We integrate patent- and firm-level data with natural language processing (NLP) tools to study firms' innovation foraging behavior.

Patent and firm data. We draw on U.S. patent data along with firm-level data on the public firms that produce them. We use granted-patent records from PatentsView for application years 1981–2018, which provide patent metadata, citation links, and CPC classifications; see Appendix A.1 for further details on sample selection and construction.

From the citation links we construct, for each patent, a (non-normalized) count of forward citations received within five years of application; this measure sits alongside three external proxies for innovation quality that we merge in: real patent value from Kogan *et al.* (2017), five-year breakthrough importance from Kelly *et al.* (2021), and creativity from Kalyani (2024). Firm-year financials, public-firm identifiers, and company names come from the CRSP/Compustat merged database, linked to patents through the CRSP `permno` supplied by Kogan *et al.* (2017). Throughout, we restrict attention to public firms with at least 50 patents over the sample window. The latter restriction ensures that for each firm we have a collection of multiple records of innovation.

Embeddings and clustering. To take the foraging model to data, we need an empirical representation of the space of ideas over which firms search. We construct one from patent text: our working assumption is that a patent's title and abstract are sufficient to describe, in words, the idea it embodies. We turn this text into geometry using a text embedding: a mapping that assigns to each patent a vector in a high-dimensional space, positioned so that patents describing similar ideas lie close together and patents describing distant ideas lie far apart. The resulting space is our empirical counterpart to the idea landscape of the model, and a firm's location in it at any date is given by the patents it has filed. Concretely, we embed the combined title and abstract of each patent using OpenAI's `text-embedding-3-small` model, yielding a 256-dimensional vector for each of roughly seven million patents.

This embedding places patents in a continuous space, but the model is organized around discrete patches—localized regions of the idea space that firms exploit and move between. We recover an empirical analogue to patches by clustering the embeddings in two stages. We first apply a density-based clustering algorithm (HDBSCAN), which lets clusters emerge from the data rather than fixing their number in advance, with parameters tuned separately within each top-level technology section. We then agglomerate the resulting groups to a common, interpretable resolution, targeting within each section the number of subclasses recorded in the patent classification system (CPC). This anchors the granularity of our clusters to an external, widely used taxonomy of technologies while still allowing their boundaries to be shaped by the text itself. Appendix A.2 provides details on the embedding and clustering procedures.[9] For patent-based analyses requiring cluster data, we typically restrict attention to patents not classified as noise and to firms with at least 50 valid non-noise clustered patents in the baseline window.[10]

Additional constructed variables. Our analysis relies on a few auxiliary variables. The first two are alternative patent-level measures of novelty, complementing Kalyani's (2024) measure of creativity. The first is commonly used in the literature (e.g., Youn *et al.*, 2015; Choi *et al.*, 2023): we construct a binary indicator according to which a patent is novel if it contains at least one CPC code combination that appears for the first time. Neither creativity nor combination-based novelty, however, can be mapped onto our foraging model. We therefore also construct a measure of novelty that is native to the space of ideas in which we embedded patents. For any patent, the *minimum distance* is the normalized embedding distance to the closest prior patent (by any firm) in a five-year backward window. Appendix Figure A.2 shows that minimum distance correlates with patent creativity and CPC-based novelty as well as breakthrough importance.[11]

To study exploration, we require a measure of firms entering and exiting patches. The binary indicator $\text{Entry}_{jct} \in \{0, 1\}$ records whether firm j enters cluster c in year t. As a baseline measure, we adopt the simplest possible definition. Firm j enters cluster c in the first year it patents there: $\text{Entry}^{FP}_{jct} = \mathbf{1}\left\{ N_{jct} > 0 \ \text{ and } \ \sum_{\tau<t} N_{jc\tau} = 0\right\}$, where N_{jct} denotes the number of patents filed by firm j in cluster c in year t. Conversely, the *exit year* of the (j, c) pair is the last year t in which

[9] As an illustration of the embedding methodology, Appendix Figure A.1 also maps selected firms' patents into a firm-specific two-dimensional embedding space and visualizes the underlying cluster structure.

[10] As a robustness check, we also consider patent classification subclasses (CPC4) as an alternative definition of idea patches. We prefer our embedding-based approach, as the classification system is endogenous to innovation: subclasses are created and split as fields fill up, and patents are retroactively reclassified under the current scheme (Lafond and Kim, 2019). CPC-based patch birth and depletion thus confound taxonomy maintenance with the idea-space dynamics we measure, whereas patent text is fixed at filing.

[11] We have experimented with a variety of alternative minimum-distance metrics: restricting attention to patents only from the same firm, relaxing the backward window, averaging over several close patents instead of the single closest patent, and more. The different measures turned out to behave very similarly, so we only present one of them.

$N_{jct} > 0$.[12]

Based on these indicators, we define two auxiliary statistics. The *duration* of a firm's activity in a cluster is defined as the number of years between entry and exit. To deal with noise in the definition of entry and exit, we typically exclude years with zero patents between entry and exit and refer to the resulting measure as *active years* or *active duration*. The *entry rate* is the number of entry events divided by the lagged number of active clusters:

$$\text{EntryRate}_{jt} = \frac{\sum_c \text{Entry}_{jct}}{\text{ActiveClusters}_{j,t-1}}, \qquad \text{ActiveClusters}_{jt} = \sum_c \mathbf{1}\left\{ N_{jct} > 0 \right\},$$

which is defined when $\text{ActiveClusters}_{j,t-1} > 0$.

Summary statistics. Table A.3 in the Appendix gives summary statistics on the resulting empirical landscape of idea-patches and firm behavior. We obtain 667 such patches, with an average of 1,223 patents per patch. Exploiting the link between a patent and its assignees, we are able to resolve which firms have visited a patch and thus obtain firm-patch level statistics. We find that the average firm's patch contains 16 patents and the average spell lasts just under six active years. The median number of patents per patch, by contrast, is only two (six when conditioning on a patch having at least two patents), and the median number of active years per spell is three.[13] Turning to (five-year) forward citations, the median number of citations to the patents it contains is around 2, while at the 25th percentile this number drops to 1 and at the 75th percentile it rises to 4.5. These statistics indicate large heterogeneity across patches in terms of both quantity and quality of patents. See Appendix A.3 for further statistics on idea patches as well as our baseline patent and firm samples.

2.2 Results

This section presents the three main facts motivating our theory.

[12] This measure is sensitive to noise in the clustering assignment process—a single patent assigned to a cluster by a firm that subsequently is not active in this cluster for several years counts as entry—so, where this concern is especially acute, we also consider an alternative "peak-flow" entry rule. Let $f_{jc,t}$ be the trailing W-year rolling mean of patent flow, $f_{jc,t} = \frac{1}{W}\sum_{k=0}^{W-1} N_{jc,t-k}$, and let $f_{jc}^{\text{peak}} = \max_t f_{jc,t}$. For a threshold $\texttt{thr} \in (0,1]$, define $\text{Entry}_{jct}^{PF} = \mathbf{1}\left\{t = \min\left\{s : f_{jc,s} \geq \texttt{thr} \cdot f_{jc}^{\text{peak}}\right\}\right\}$. We set $W = 3$ and $\texttt{thr} = 0.1$. The definition of exit is symmetric. Conditional on entry and exit by this rule, cluster-year activity requires only a single patent per year.

[13] These statistics use the baseline sample, which includes spells censored by the 1981–2018 window. The completed spells Fact 2 is based on are shorter, with the median being a single active year, as long spells are the most likely to be censored.

2.2.1 Fact 1: Diminishing returns to exploitation at the patch level

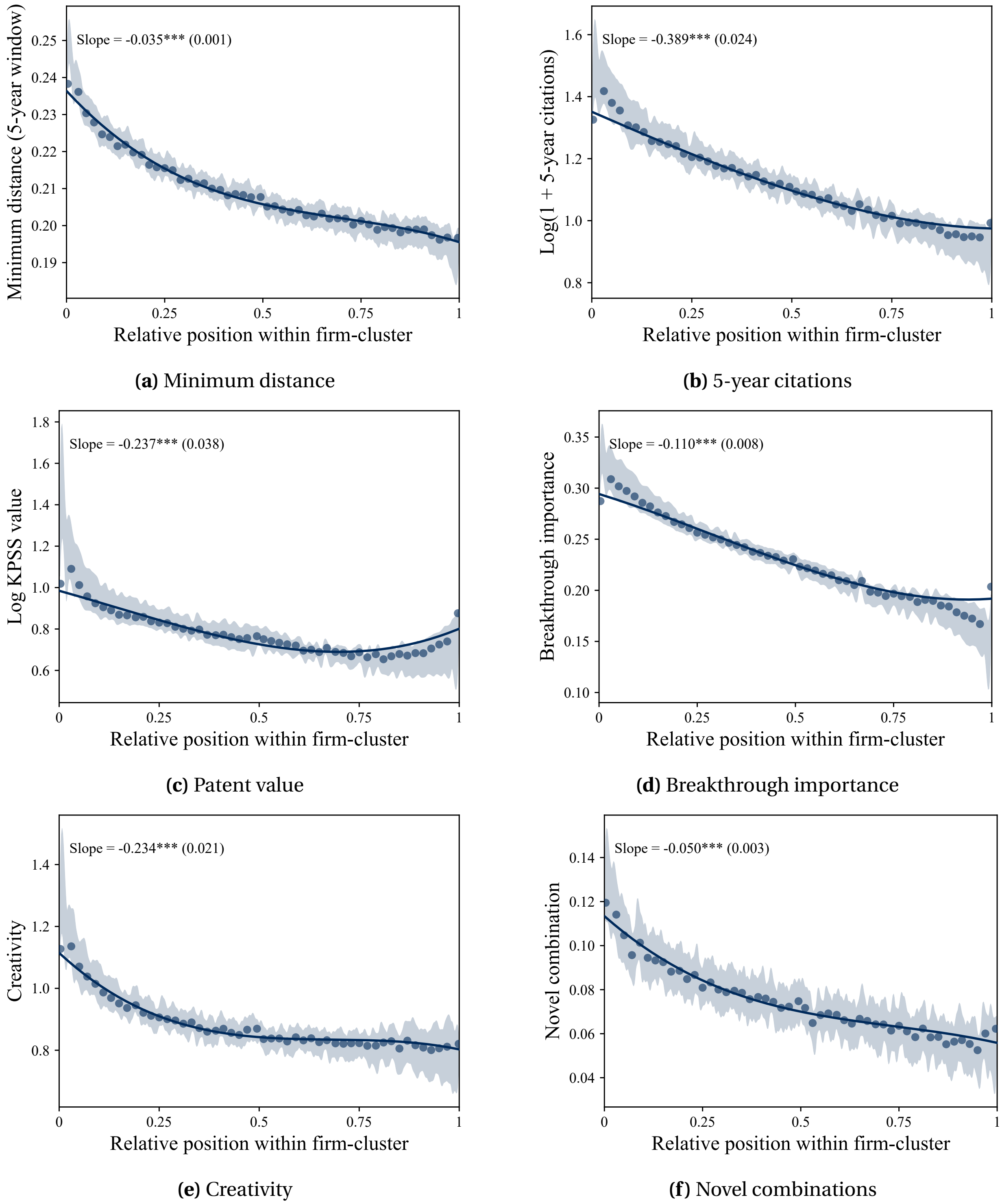
We begin with the within-patch dimension of foraging: how the quality of a firm's innovations evolves as it continues to exploit a given idea patch. The direction is not obvious a priori. If a patch holds a finite stock of valuable ideas, the best of which are found first, then later patents should embody progressively smaller advances as the patch is drawn down. A learning-by-doing logic points the other way: as a firm accumulates experience within a patch, it may grow more proficient at extracting ideas from it, so that quality rises with familiarity. Which force dominates is an empirical question, and this subsection takes it up.

Figure 2 relates six measures of patent quality to a patent's relative position within its firm-cluster sequence, our empirical counterpart to a firm's progression through an idea patch. Thus, for each firm and cluster, we order patents by application date and assign a relative position on a zero-to-one scale, so that the first patent in a sequence takes the value zero and the last takes the value one; firm-clusters containing a single patent are excluded. The outcomes comprise the five-year forward citation count and the three external proxies introduced above—patent value, breakthrough importance, and creativity—together with the embedding-based minimum distance and the classification-based indicator for novelty. Each panel plots binned means with a cubic fit, after absorbing application-year and firm-by-cluster fixed effects, with confidence bands constructed from standard errors clustered by cluster. For each panel, we also report the estimates of a simple linear model.

The six panels share a common pattern: along every measure, quality declines as the firm advances through the patch. The patents a firm files late in a patch are, on average, less novel, less cited, less valuable, less important, and less creative than the ones it filed first.[14] We read this as decreasing returns to exploitation within an idea patch: as a firm accumulates patents in a region of the idea space, the marginal patent it produces there embodies a smaller advance than the one before, as though the most productive ideas in the patch are drawn down first and the patch is progressively depleted of the improvements still available to be found.

This is not a reflection of declining effort. Appendix A.4.2 shows that the number of inventors per patent, a simple proxy for research input, does not decline over the sequence, so successive patents represent at least comparable input for a diminishing output. Further, these patterns are robust to how idea patches are delineated. Appendix A.4.1 reproduces the figure under alternative cluster definitions—global rather than firm-specific embedding clusters, and clusters built from the patent classification system—and obtains the same decline across measures.

[14] The patent-value measure should be interpreted with caution. In this context, unlike the other measures, it does not survive firm-by-year fixed effects. This reflects the construction of the KPSS measure, which assigns the firm's grant-day stock return to the patents granted that day. In general, much of the variation in the KPSS measure is therefore firm-level rather than patent-specific.



(a) Minimum distance

(b) 5-year citations

(c) Patent value

(d) Breakthrough importance

(e) Creativity

(f) Novel combinations

Figure 2: Relative position and innovation outcomes

Notes. Relative position orders patents within firm-cluster sequences after excluding singleton firm-clusters and scales each sequence from 0 to 1. Panels plot binned means with a cubic fit. Importance and creativity use 1981–2016; the other panels use 1981–2018.

Table 1: Early patents and top innovation outcomes

Outcome	Top decile	Middle 45–55%	Bottom decile
Panel A: first 10 patents share			
Minimum distance (5-year)	0.299	0.234	0.192
Citations	0.322	0.209	0.180
Patent value	0.315	0.220	0.246
Importance	0.405	0.216	0.176
Creativity	0.346	0.239	0.187
Novelty	0.404	–	0.227
Panel B: first 10% patents share			
Minimum distance (5-year)	0.151	0.114	0.074
Citations	0.203	0.095	0.068
Patent value	0.211	0.089	0.126
Importance	0.301	0.089	0.057
Creativity	0.197	0.110	0.071
Novelty	0.165	–	0.110

Notes. The sample is the baseline restricted to firm-clusters whose first observed patent and last observed patent in the full patent-level file both fall within 1981–2018 and that contain more than 10 patents in the table sample. Outcome buckets are firm-internal percentile buckets. Novelty is binary: its top column reports novel-combination patents, its bottom column reports non-novel patents, and the middle column is not defined. Each cell reports the share of patents in that outcome bucket that are early within their firm-cluster. Patent value is ranked in raw real values.

Finally, we also show that this pattern holds within different broad technology areas (Appendix Figure A.4): a negative relation between each of the quality/novelty measures and the order of patents for a given firm-patch obtains across different technology sections.

Table 1 gives a sense of the magnitudes involved by asking a simple question: of the highest-quality patents a firm produces in a patch, what fraction does it produce early? Among a firm's patents in a patch, patents in the most-cited decile are filed earlier than the rest. Of these top-cited patents, 20.3% appear in the opening tenth of the patch; of its median-cited patents, only 9.5% do (those of typical quality for the firm and patch, in the 45th–55th percentile of its within-patch citation distribution). The gap is wider for breakthrough importance (30.1% versus 8.9%) and comparable for patent value (21.1% versus 8.9%). Defining the early window as a firm's first ten patents in a patch, rather than its first tenth, leaves the ordering unchanged.[15]

[15] This exercise is descriptive and uses a different sample from Figure 2: it conditions on completed firm-patch spells and on patches with more than ten patents, and reports raw shares without fixed effects. These restrictions follow from the statistic itself: an "early share" is defined only for a completed spell, and ranking patents within a patch requires enough of them to populate the deciles. Novel combination is binary, so its middle group is undefined.

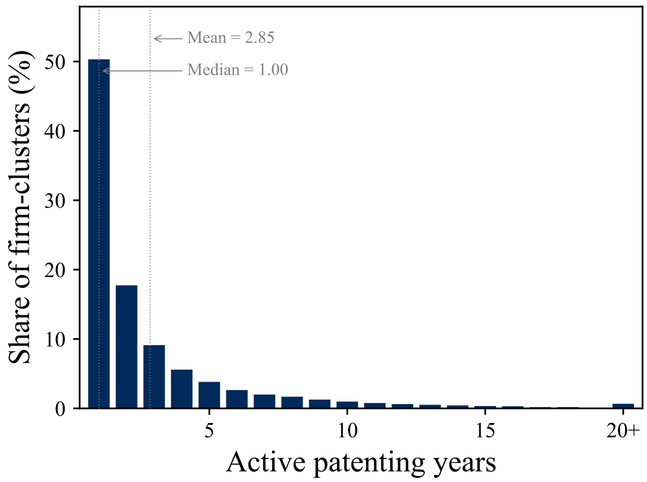


Figure 3: Distribution of cluster durations

Notes. The figure uses firm-cluster spells whose first patent entry and last patent exit both fall within 1981–2018. Duration is measured as active patenting years.

Interpretation. Taken together, Figure 2 and Table 1 establish that a firm's innovations within a patch deteriorate, on every measure of quality, as it continues to exploit that patch. This points to some form of decreasing returns operating at the level of the individual idea patch. It is suggestive of the broader notion that ideas are becoming harder to find (Bloom *et al.*, 2020), but locates that difficulty within the firm's own trajectory through a patch: the same firm, working the same region of the idea space, finds successively smaller improvements the longer it stays. This is, however, only one side of foraging. It describes what happens while a firm exploits a patch, and says nothing yet about when it chooses to leave.

2.2.2 Fact 2: The dynamics of exploration across patches

The decline documented in Figure 2 is not a technological fate that firms passively absorb. A firm facing diminishing returns within a patch can respond: it can stop working that patch and begin a new one elsewhere in the idea space. This is the second element of foraging, the movement across patches, and it is the margin we turn to now—how often firms enter new patches, and how long they stay.[16] In the data, firms enter new patches routinely. In a typical firm-year, the number of new patches a firm enters is a substantial fraction of the patches it was already exploiting. These entries differ greatly in what they become. Many lead nowhere: the firm files one or two patents in the new patch and moves on. A minority develop into extended runs of patenting. Figure 3 shows this by plotting the distribution of spell durations, measured as the number of

[16]Because a firm leaves a patch at a time of its choosing, the firm-cluster sequences behind Fact 1 are truncated: we do not observe how far quality would have fallen had the firm continued. The within-patch decline we measure is, in this sense, a lower bound.

Table 2: Cluster quality and duration

	Share active years		
	(1)	(2)	(3)
Log average 5-year citations	0.0322***		
	(0.0013)		
Average breakthrough importance		0.0202***	
		(0.0057)	
Log average patent value			0.0214***
			(0.0013)
Observations	28,433	25,654	28,433
R^2	0.284	0.250	0.266

Notes. The table reports regressions at the firm-cluster level. The outcome is the firm-cluster share of the firm's active patenting years. The sample keeps firm-clusters whose first-entry year is at least 1981 and whose last observed patent year is no later than 2018. Patent-level quality measures are winsorized at p1/p99 before being averaged to the firm-cluster level. Each regression includes firm fixed effects. Standard errors are clustered at the firm-level.

active patenting years between a firm's first and last patent in a patch. The distribution is heavily right-skewed: the median completed spell lasts a single active patenting year, and about half consist of a single patent. The remainder are more substantial—among completed spells with at least two patents, the median runs to three active patenting years over a calendar span of eight, and the longest spells last for decades. How long a firm stays in a patch thus varies widely. This variation suggests that the length of a spell is not dictated by the patch alone, but reflects a choice about when to leave it—a choice that should respond to how much the patch still has to offer, which we examine next.

Patch quality and duration. If firms choose when to leave a patch, the length of a spell should reflect the quality of the patch and not technology alone: a firm should remain longer on a richer patch and leave a poorer one sooner. We find that it does. We measure how much of its patenting a firm devotes to a patch by the active patenting years it records there—the number of distinct years in which it files at least one patent in the patch—taken as a share of the active patenting years it records across all of its patches. This share captures how a firm divides its patenting time across the patches it works. We relate it to three empirical proxies for patch quality: average forward citations, breakthrough importance, and patent value. Table 2 shows that all three are positively related to a patch's share of firm active patenting years. Firms spend proportionately more of their patenting time on richer patches, consistent with the stopping rule we develop below.

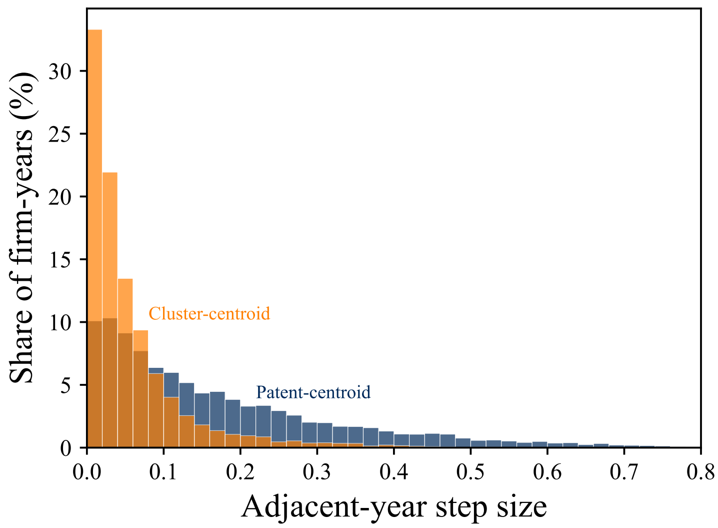


Figure 4: Step size distribution at the firm-year level

Notes. The figure uses adjacent active firm-years in the public-firm baseline cluster sample. The patent-centroid series uses annual patent-embedding averages, and the active-cluster-centroid series uses annual active-cluster centroid averages. Step size is measured as cosine distance between adjacent-year positions.

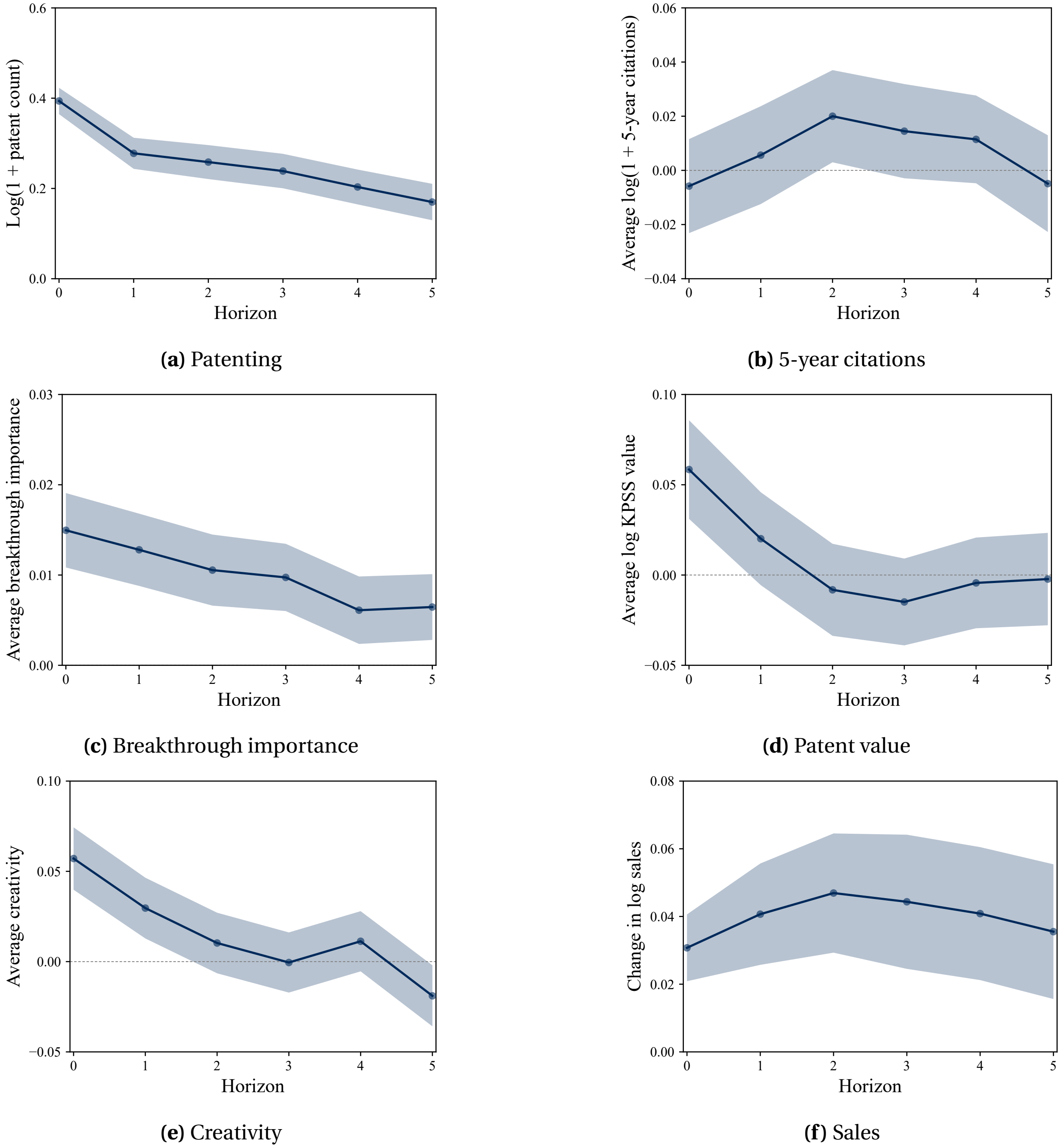
2.2.3 Fact 3: Firm-level movements and outcomes reflect this foraging

The preceding sections characterized intra- and inter-patch dynamics at the patent level. We now aggregate and document how foraging for ideas expresses itself at the firm-year level: through the distribution of steps that firms take through technology space, and the dynamic payoffs to entering new clusters.

Steps through technology space. We summarize a firm's movement through idea space by the step-size distribution at the firm-year level. We define a firm's position in each year by averaging either across all patent embeddings filed that year or across the centroids of its active clusters. The step size is the cosine distance between adjacent-year positions. Figure 4 shows this distribution is distinctly right-tailed: small steps in most years are interspersed with much longer ones. Apple's trajectory in Figure 1c is not the exception but the rule.

Dynamic payoffs to cluster entry. Do firms benefit from these occasional large steps, from entering new clusters? We trace the dynamic association between new-cluster entry and subsequent firm outcomes using local projections. Let Y^m_{jt} denote a firm-year outcome: for patent-quality measures, Y^m_{jt} is the average of patent-level measure m across patents filed by firm j in year t; for accounting outcomes, Y^m_{jt} is log sales. For each outcome m and horizon $h = 0, \ldots, 5$, we estimate

$$\Delta_h Y^m_{jt} = \alpha^m_j + \delta^m_t + \beta^m_h \text{EntryRate}_{jt} + X'_{j,t-1}\Gamma^m_h + \varepsilon^{m,h}_{jt},$$



(a) Patenting

(b) 5-year citations

(c) Breakthrough importance

(d) Patent value

(e) Creativity

(f) Sales

Figure 5: Local projections around first-patent cluster entry

Notes. The treatment is the first-patent entry rate. Horizons run from $h = 0$ to $h = 5$. Patent outcomes are measured at $t + h$; firm outcomes are cumulative changes from $t - 1$ to $t + h$. Regressions include firm and treatment-year fixed effects, lagged firm controls, and firm-clustered standard errors. The sample requires t, $t - 1$, and $t + h$ to fall within 1981–2018.

where

$$\Delta_h Y_{jt}^m = \begin{cases} Y_{j,t+h}^m, & \text{for patenting and patent-quality outcomes,} \\ Y_{j,t+h}^m - Y_{j,t-1}^m, & \text{for sales.} \end{cases}$$

The "treatment" is the first-patent cluster entry rate. Controls $X_{j,t-1}$ include lagged log sales, lagged cumulative patenting, and ActiveClusters$_{j,t-1}$. We include firm and treatment-year fixed effects, cluster standard errors at the firm level, and require t, $t-1$, and $t+h$ to lie in 1981–2018. The coefficients β_h^m trace the dynamic response of firm outcomes to new-cluster entry.

Figure 5 displays the results. Cluster entry is associated with an immediate increase in patent counts, which then gradually fades. The average quality of patents rises and likewise diminishes as the horizon extends. Sales are elevated for several years. In sum, entering new clusters is associated with firm-wide increases in the quantity and quality of patents and, downstream, in sales.

3 Theoretical framework

Guided by the facts established in the preceding section, this section develops our theoretical framework. At its core is a firm that can either exploit a patch of related ideas, improving the quality of its active product line through steps that diminish in size as the patch is depleted, or explore the idea space for a new patch of unknown quality. We embed this foraging problem in a general-equilibrium endogenous growth model. Section 3.1 describes the environment, Section 3.2 derives the balanced-growth path equilibrium, and Section 3.3 offers analytical characterizations of key equilibrium objects.

3.1 Environment

Time is continuous and runs forever. We consider the economy in a balanced-growth-path equilibrium.

3.1.1 Households and firms

Representative household and final consumption. The economy admits an infinitely-lived representative household with log utility over consumption, discounted at a rate ρ: $U = \int_0^\infty e^{-\rho t} \log(C_t)\, dt$. The household inelastically supplies L units of labor and owns all firms.

The final consumption good is produced by a competitive sector that aggregates a continuum of differentiated varieties $\mathcal{N}(t)$ using a CES technology with elasticity of substitution $\sigma > 1$

$$Y = \left(\int_0^{\mathcal{N}(t)} \left[q(j)\, y(j) \right]^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

where $q(j)$ is the quality of variety j and $y(j)$ is the quantity.

Variety production. Each variety is produced by a monopolist with a linear, labor-only technology $y(j) = \ell(j)$, the cost of labor being wage w. Defining the aggregate quality index $Q \equiv \left(\int_0^{\mathcal{N}} q_j^{\sigma-1}\, dj \right)^{1/(\sigma-1)}$, and with the final good as numeraire, this implies economy-wide output, wage, and profit as a function of quality:

$$Y = Q \cdot L^P, \qquad w = \frac{\sigma - 1}{\sigma} Q, \qquad \pi(q) = \kappa\, q^{\sigma-1} \tag{2}$$

where L^P is total production labor and $\kappa \equiv \frac{L^P}{\sigma} Q^{2-\sigma}$ is profit-per-unit-of-quality. It is convenient to define "economic quality" $\tilde{q} \equiv q^{\sigma-1}$, as all flow payoffs are linear in $\tilde{q}$, e.g., $\pi(\tilde{q}) = \kappa\, \tilde{q}$.

Firms. There is an endogenous mass F of firms. A firm is a collection of product lines, each with quality q_j. At any time, a firm has one *active* product line with current quality q, which is the focus of the firm's R&D; and a *legacy* portfolio $\mathcal{L} = \{q_1, \ldots, q_n\}$ of products that are produced and sold but no longer improved. The process of product addition and improvement is the central focus of the model and is described in detail below. We keep exit simple. Each individual product line (active or legacy) faces independent creative destruction at Poisson rate δ. Additionally, each firm faces an exogenous death shock at Poisson rate $\delta_F > 0$, independent across firms and independent of line-level creative destruction. Death destroys the firm and all its products. The effective hazard of losing any given product line is $\delta + \delta_F$. When the active product is destroyed, the product is lost and the firm enters exploration to find a new product line.

There is free entry. A potential entrant pays a sunk cost of K_R units of labor. It enters with an empty legacy portfolio and starts exploiting a patch of randomly drawn type to improve a new product of quality $q_0 = \bar{\Lambda}_E \cdot \bar{Q}_t$, where $\bar{Q}_t$ is the *average* product quality in the economy:

$$\bar{Q}_t \equiv \left(\frac{1}{\mathcal{N}_t} \int_0^{\mathcal{N}_t} q_j^{\sigma-1}\, dj \right)^{1/(\sigma-1)} \qquad \leftrightarrow \qquad Q_t = \mathcal{N}_t^{1/(\sigma-1)}\, \bar{Q}_t. \tag{3}$$

3.1.2 Innovation as foraging: exploitation and exploration

At the core of the model stands the firm foraging for ideas. We model each firm as pursuing a composite search strategy, alternating between exploration and exploitation, with switching between both modes optimally and endogenously responding to the diminishing potential of a patch.

Patches, foraging, and patch dynamics. Each firm moves in a space of well-defined *patches*, each consisting of a set of closely related ideas. We assume the number of patches is arbitrarily large, so there is no re-encounter. We furthermore suppose that the patch of ideas corresponds to, or applies to, one product line. Importantly, each patch is characterized by a state X, the "density" of the patch, which tracks how much more room for improvement of the product's quality there is. Patches are potentially heterogeneous in their initial density $x_{0,k}$, where the share of patches belonging to type k is p_k, where $\sum_{k=1}^{K} p_k = 1$ and $K \in \mathbb{Z}_{++}$. The forager either *exploits* a patch for its ideas, depleting it in the process, or *explores* the landscape to find a new patch.

Exploitation. While on a patch, patents arrive according to a homogeneous Poisson process with constant intensity γ, each improving active-product quality multiplicatively: $q \rightarrow \theta(X)\,q,\ \theta(X) > 1,\ \theta'(X) > 0$. This captures the central assumption that step sizes are *increasing* in patch density.

During exploitation, the density of that patch, X, evolves according to a Brownian motion with negative drift,

$$dX(t) = \mu dt + \sigma_x dW_x(t), \tag{4}$$

where $\mu < 0$, the noise parameter is σ_x, and $X(0) = x_{0,k}$. The negative drift $\mu < 0$ captures the gradual exhaustion of opportunities for further quality improvements ("destructive foraging" in the language of behavioral ecology).[17]

Exploration. When the firm abandons a patch, the active product ($\tilde{q}_{\text{exit}}$) becomes legacy, earning $\kappa\tilde{q}_{\text{exit}}$ until destroyed, and the firm's R&D involves exploring the landscape in a random direction until, at Poisson rate λ, it discovers a new patch. The quality of the new product line scales proportionally with the firm's frontier quality at exit, $q_{\text{new}} = \Lambda q_{\text{exit}}$, where Λ captures the firm's loss in capability when moving to a new cluster.

[17] We model depletion of patch density as occurring at a fixed rate μ, i.e., it does not vary with the actual rewards taken out of the patch. One interpretation of an exogenous μ is that it reflects other foragers' activity, either in the given patch or elsewhere, making the ideas in the given patch less valuable. We could alternatively model depletion as reward-based, where each arrival of a reward causes X to jump down by an increment that is increasing in the magnitude of the reward.

R&D costs. R&D costs in both exploitation and exploration scale with (relative) quality. This assumption is analogous to the standard assumption in Klette and Kortum (2004) and captures the notion that innovation on more complex products requires proportionally more resources.

Let z denote the *average-quality weight*

$$z \equiv \left(q/\bar{Q}\right)^{\sigma-1}. \tag{5}$$

R&D costs scale with the aggregate-quality share $s = (q/Q)^{\sigma-1} = z/\mathcal{N}$. R&D labor in exploitation is $\ell^I = \frac{c_I}{\sigma-1}\frac{z}{\mathcal{N}}$, with flow cost $w \cdot \ell^I = \frac{c_I}{\sigma} Q^{2-\sigma} q^{\sigma-1} = \hat{c}_I \kappa q^{\sigma-1}$, where $\hat{c}_I \equiv c_I/L^P$. Similarly for exploration:

$$\ell^E = \frac{c_E}{\sigma-1}\frac{z_{\text{ref}}}{\mathcal{N}}, \tag{6}$$

where $z_{\text{ref}} = (q_{\text{ref}}/\bar{Q})^{\sigma-1}$ is the reference quality weight from the last exploitation spell.

Spatial overlay. In line with the foraging literature, the two modes of innovation correspond to different *movement* patterns. It bears emphasizing that this is a pure measurement layer; the single economic state variable remains the scalar patch density X.

Exploitation, or intensive-mode search, is associated with local diffusion. The forager's position $Z(t) \in \mathbb{R}^2$ evolves according to $dZ(t) = \sigma_I \, dW_z(t)$, where $W_z(t)$ is a standard two-dimensional Brownian motion and $\sigma_I > 0$ governs the scale of within-patch movement. The process is initialized at the patch center upon entry. If T_n denotes the arrival time of the nth patent, its observed location is $Z(T_n)$.

By contrast, during *exploration*, or extensive-mode search, the firm is looking for a new patch by drawing a direction uniformly on the circle and traveling ballistically at speed v_E until the next patch arrives. Because the direction is fixed within a spell, the path is a straight ray, and the distance covered equals v_E times the time spent searching. [18]

3.2 Balanced growth path equilibrium

We characterize a balanced-growth path (BGP) equilibrium on which the aggregate quality index Q_t, and hence consumption, grow at a common rate g.

[18]Exploration ends at the first of two independent Poisson events, patch arrival (rate λ) and firm death (rate δ_F), so conditional on arriving before death the search time is exponential with rate $\lambda + \delta_F$. Scaling by the constant speed v_E, the distance traveled to the new patch is exponential: $f(d \mid \text{entry}) = \alpha_d \, e^{-\alpha_d d}, \ \alpha_d \equiv \frac{\lambda+\delta_F}{v_E}$.

3.2.1 Value functions

Let $\mathcal{A}(\tilde{q}, X, t)$ denote the value, in final-good units, of an active product with economic quality $\tilde{q}$ on a patch with density X. Let $\mathcal{A}^E(\tilde{q}, t)$ denote the value of searching for a new patch after leaving an active line with reference quality $\tilde{q}$.

The key property underpinning tractability is homogeneity in economic quality. Flow profits, R&D costs, patent improvements, and new-patch discoveries all scale proportionally with $\tilde{q}$. Hence the level of product quality can be factored out of the firm problem. This homogeneity property hinges on four conditions: (i) CES demand (so profits are linear in $\tilde{q}$); (ii) multiplicative quality improvements; (iii) R&D costs proportional to $\tilde{q}$; and (iv) new-product quality proportional to abandoned-patch product quality.

On a BGP, the remaining aggregate time dependence comes through the profit shifter $\kappa(t)$, so we write

$$\mathcal{A}(\tilde{q}, X, t) = \kappa(t)\tilde{q}\,\hat{W}(X), \qquad \mathcal{A}^E(\tilde{q}, t) = \kappa(t)\tilde{q}\,\hat{w}^E. \tag{7}$$

Household optimization yields the standard Euler equation $r = \rho + g$, where r is the interest rate. Since $\kappa(t) = \frac{L^P}{\sigma} Q_t^{2-\sigma}$ and Q_t grows at rate g, $\dot{\kappa}/\kappa = (2-\sigma)g$. Thus normalized values are discounted at

$$\hat{\rho} \equiv r - \frac{\dot{\kappa}}{\kappa} = \rho + (\sigma - 1)g. \tag{8}$$

After this normalization, neither calendar time nor product quality remains as a state variable.

Given this structure, we can characterize the firm's dynamic choice about when to abandon the current patch and begin searching for a new one as a one-dimensional optimal stopping problem (e.g., Stokey, 2008). Conditional on a candidate BGP pair (g, L^P), the abandonment decision depends only on the current patch density X. Higher X raises the value of continuing through larger expected quality improvements, while the value of abandonment, $\hat{W}_0 + \hat{w}^E$, does not depend on X. The optimal stopping policy is therefore summarized by a *threshold* in X, denoted x^* (monotonicity of $\hat{W}$ in X follows from $\theta' > 0$ and the additive dynamics of X; we verify it in all numerical solutions): the firm exploits the current patch when $X > x^*$ and abandons it when $X \leq x^*$. We next characterize this threshold.

Legacy value. Legacy products are passive: they continue to earn profits but are no longer improved. The firm may own a portfolio of such products, but their values are additive and unaffected by the decision to continue exploiting the current patch or abandon it. Hence the stopping problem can be written for the active line alone. When the firm abandons the active line, that line becomes a legacy product, whose normalized value is $\hat{W}_0$. A legacy product earns flow profit $\kappa\tilde{q}$ and is destroyed either by product-level creative destruction, at rate δ, or by firm death, at rate δ_F. Its normalized value is

$$\hat{W}_0 = \frac{1}{\hat{\rho} + \delta + \delta_F}. \tag{9}$$

Exploitation. During exploitation, the active product yields net flow profit $(1 - \hat{c}_I)$ per unit of economic quality. In the continuation region $X > x^*$, patents arrive at rate γ, creative destruction arrives at rate δ, and firm death arrives at rate δ_F. Voluntary abandonment enters through the boundary conditions at x^*. The normalized value therefore is

$$(\hat{\rho} + \delta + \delta_F)\,\hat{W} = (1 - \hat{c}_I) + \gamma\left[\theta(X)^{\sigma-1} - 1\right]\hat{W} + \delta\hat{w}^E + \mu\,\hat{W}' + \tfrac{\sigma_x^2}{2}\,\hat{W}''. \tag{10}$$

Together with value matching and smooth pasting below, this second-order ODE is completed by an upper-tail condition that selects the non-explosive solution.[19]

Derivation. To derive equation (10), note that in the continuation region, $dX = \mu dt + \sigma_x dW_x$. Ito's lemma gives $\mathbb{E}[d\hat{W}(X)] = \left(\mu\hat{W}'(X) + \frac{\sigma_x^2}{2}\hat{W}''(X)\right) dt$. Over the same interval, the firm earns flow payoff $(1 - \hat{c}_I)dt$; a patent arrives with probability γdt and raises normalized value by $\theta(X)^{\sigma-1}\hat{W}(X) - \hat{W}(X)$; creative destruction arrives with probability δdt and gives continuation $\hat{w}^E$; and firm death arrives with probability $\delta_F dt$ and gives zero. Hence

$$\hat{\rho}\hat{W} = (1 - \hat{c}_I) + \mu\hat{W}' + \frac{\sigma_x^2}{2}\hat{W}'' + \gamma[\theta^{\sigma-1} - 1]\hat{W} + \delta(\hat{w}^E - \hat{W}) - \delta_F\hat{W}.$$

Rearranging yields (10).

Exploration. To characterize the (normalized) continuation value $\hat{w}^E$, note that during exploration the firm pays the R&D flow cost $\hat{c}_E$ per unit of (reference) economic quality $\tilde{q}$ and either finds a new patch (rate λ) or dies exogenously (rate δ_F). Hence,

$$(\hat{\rho} + \lambda + \delta_F)\,\hat{w}^E = -\hat{c}_E + \lambda\Lambda^{\sigma-1}\sum_{k=1}^{K} p_k\,\hat{W}(x_{0,k}). \tag{11}$$

Optimal stopping. Value matching implies that at the threshold x^*, the firm is indifferent between continuing and switching.

$$\hat{W}(x^*) = \hat{W}_0 + \hat{w}^E = \frac{1}{\hat{\rho} + \delta + \delta_F} + \hat{w}^E. \tag{12}$$

[19] When $\theta(\cdot)$ is bounded, this is equivalent to requiring $\hat{W}(X)$ to remain bounded as $X \to \infty$.

So, at x^*, the value of the active product including continuation equals the value of shelving ($\hat{W}_0$) plus searching for the next product ($\hat{w}^E$).

Moreover, as the abandonment value is independent of X, smooth pasting gives

$$\hat{W}'(x^*) = 0, \tag{13}$$

so a marginal delay in switching yields zero net benefit.

3.2.2 Firm-level growth rate

Before turning to the aggregate equilibrium, it is instructive to derive the average rate at which a firm accumulates log quality over its innovation cycle, $g_q \equiv \frac{E[G]}{E[T]}$, where $E[G]$ is the expected log-quality gain over a cycle and $E[T]$ is the expected cycle length, with firm death cutting a cycle short. In steady state, g_q is the flow of log-quality gains per firm per unit of time; when $\delta_F = 0$, it is also, by the renewal–reward theorem, the long-run growth rate of an individual firm's log quality.

A cycle begins with exploitation of a patch. The expected duration of an exploitation spell starting from patch density X is denoted $m(X)$, and the expected log-quality gain accumulated through patents during that spell is $h(X)$. By the Feynman–Kac formula, which converts expected values of functionals of a diffusion into deterministic ODEs, these objects solve the following boundary value problems (BVPs):

$$\frac{\sigma_x^2}{2}\, m'' + \mu\, m' - (\delta + \delta_F)\, m + 1 = 0, \qquad m(x^*) = 0,\ m \text{ bounded}, \tag{14}$$

$$\frac{\sigma_x^2}{2}\, h'' + \mu\, h' - (\delta + \delta_F)\, h + \gamma \log \theta(X) = 0, \qquad h(x^*) = 0,\ h \text{ bounded}. \tag{15}$$

Averaging over initial patch types gives $\bar{\tau}^I = \sum_k p_k m(x_{0,k})$ and $\bar{h} = \sum_k p_k h(x_{0,k})$.

The firm reaches exploration unless it dies during exploitation. Since firm death arrives at rate δ_F, the probability of death during the exploitation spell is $p_{\text{death},I} = \delta_F \bar{\tau}^I$. Conditional on reaching exploration, the search spell ends at the first of discovery, at rate λ, or firm death, at rate δ_F. Hence its expected duration is $1/(\lambda + \delta_F)$, and discovery occurs before death with probability $\lambda/(\lambda + \delta_F)$. It follows that expected log-quality gain and expected cycle time are

$$E[G] = \bar{h} + (1 - \delta_F \bar{\tau}^I)\frac{\lambda}{\lambda + \delta_F} \log \Lambda, \tag{16}$$

$$E[T] = \bar{\tau}^I + \frac{1 - \delta_F \bar{\tau}^I}{\lambda + \delta_F} \equiv \bar{T}. \tag{17}$$

Therefore,

$$g_q = \frac{E[G]}{E[T]} = \frac{\bar{h} + (1 - \delta_F \bar{\tau}^I)\frac{\lambda}{\lambda+\delta_F}\log\Lambda}{\bar{\tau}^I + \frac{1-\delta_F\bar{\tau}^I}{\lambda+\delta_F}}. \tag{18}$$

When $\Lambda = 1$, this reduces to

$$g_q = \frac{\bar{h}}{\bar{\tau}^I + \frac{1-\delta_F\bar{\tau}^I}{\lambda+\delta_F}}.$$

Thus, in this special case, g_q is the expected log-quality gain from patents during exploitation divided by expected cycle length. This object will be useful for interpreting the model's firm-level mechanism. Aggregate growth g is pinned down separately below.

3.2.3 General-equilibrium conditions

The remaining conditions close the model in aggregate. Renewal accounting pins down firm composition and the product-to-firm ratio. The stationary quality distribution then determines aggregate growth, R&D labor demand, and free entry.

Stationary firm composition and product varieties. Let ζ denote the steady-state flow of firms into exploitation. By Little's law,

$$F^I = \zeta\bar{\tau}^I, \qquad F^E = \zeta\frac{1-\delta_F\bar{\tau}^I}{\lambda+\delta_F}, \qquad F = F^I + F^E. \tag{19}$$

Using (17), $\zeta = F/\bar{T}$ and the stationary firm shares are

$$f_I \equiv \frac{F^I}{F} = \frac{\bar{\tau}^I}{\bar{T}}, \qquad f_E \equiv \frac{F^E}{F} = \frac{1-\delta_F\bar{\tau}^I}{(\lambda+\delta_F)\bar{T}}. \tag{20}$$

Entry replaces firm death in steady state,

$$E = \delta_F F. \tag{21}$$

Each variety faces destruction at rate δ or δ_F, independently. New varieties come from incumbents completing exploration and from entrants:

$$\dot{\mathcal{N}} = \lambda F^E + E - (\delta + \delta_F)\mathcal{N}. \tag{22}$$

In steady state, $\dot{\mathcal{N}} = 0$, so $\mathcal{N}^* = \frac{\lambda F^E + E}{\delta+\delta_F}$.

To convert per-firm objects into per-variety objects it is useful to define the steady-state product-to-firm ratio, using $E = \delta_F F$,

$$\chi_{\mathcal{N}} \equiv \frac{\mathcal{N}}{F} = \frac{\lambda f_E + \delta_F}{\delta + \delta_F}. \tag{23}$$

Average-quality weights. Recall from (5) that the average-quality weight of product j is $z_j \equiv \left(\frac{q_j}{\bar{Q}}\right)^{\sigma-1}$. By the definition of $\bar{Q}$, the cross-sectional mean of z_j across products equals one on a BGP. The aggregate conditions below require three stationary per-firm averages. These are the weight at exploitation entry, $\bar{z}_{\text{entry}}$; the average weight of exploiting firms, $\bar{z}_I$; and the average reference weight of exploring firms, $\bar{z}_E$.

Two within-spell objects summarize the required quality-weight accounting. Let M_t denote the spell-level quality-weight multiplier, net of aggregate dilution. Define $\bar{\Psi}^{\text{surv}}$ as the expected terminal multiplier at non-death exit, counting zero if the firm dies during exploitation. Let $\bar{\Xi}$ be the expected cumulative quality weight generated during exploitation. Appendix B.1 derives these objects from Feynman–Kac equations. Given them, the stationary averages are

$$\bar{z}_{\text{entry}} = \frac{\delta_F \bar{T}\, \bar{\Lambda}_E^{\sigma-1}}{1 - \dfrac{\lambda\, \Lambda^{\sigma-1}\, \bar{\Psi}^{\text{surv}}}{\lambda + \delta_F + (\sigma-1)g}}, \tag{24}$$

$$\bar{z}_I = \bar{z}_{\text{entry}} \frac{\bar{\Xi}}{\bar{\tau}^I}, \tag{25}$$

$$\bar{z}_E = \bar{z}_{\text{entry}} \frac{\bar{\Psi}^{\text{surv}}}{1 - \delta_F \bar{\tau}^I} \frac{\lambda + \delta_F}{\lambda + \delta_F + (\sigma-1)g}. \tag{26}$$

A stationary average-quality distribution requires the contractivity condition

$$\frac{\lambda\, \Lambda^{\sigma-1}\, \bar{\Psi}^{\text{surv}}}{\lambda + \delta_F + (\sigma-1)g} < 1. \tag{27}$$

This condition says that the expected quality-weight multiplication of a surviving incumbent cycle must be dominated by death and aggregate dilution.

Aggregate growth. Aggregate growth is then pinned down by stationarity of average-quality weights. Since $\bar{Q}$ is average product quality, stationarity requires $\frac{1}{\mathcal{N}} \int_0^{\mathcal{N}} z_j \, dj = 1$. Equivalently, the total z-weight of active and legacy products must equal $\mathcal{N}$. Appendix B.2 shows that this condition is

$$\frac{\bar{z}_{\text{entry}}}{\chi_{\mathcal{N}} \bar{T}} \frac{[(\sigma-1)g + \delta_F]\bar{\Xi} + \bar{\Psi}^{\text{surv}}}{(\sigma-1)g + \delta + \delta_F} = 1. \tag{28}$$

If g is too low, firms' quality weights drift upward relative to $\bar{Q}$; if g is too high, aggregate dilution is too strong. Equation (28) selects the growth rate that keeps the cross-sectional average-quality-weight distribution stationary.

The same equilibrium condition has a useful flow interpretation. Let $\Phi_z(x)$ denote the stationary density of active-product quality weight across patch densities (see Appendix B.3 for details). Then

$$(\sigma-1)g = \underbrace{\frac{1}{\mathcal{N}}\int \gamma\left[\theta(x)^{\sigma-1}-1\right]\Phi_z(x)\,dx}_{\text{patenting/quality improvement}} + \underbrace{\lambda\Lambda^{\sigma-1}\frac{f_E}{\chi_{\mathcal{N}}}\bar{z}_E}_{\text{discovery}} \underbrace{-\,\delta}_{\text{obsolescence}} + \underbrace{\delta_F\left(\frac{\bar{\Lambda}_E^{\sigma-1}}{\chi_{\mathcal{N}}}-1\right)}_{\text{net entry}}. \quad (29)$$

The first term is quality improvement from patents on active products. The second is quality creation when exploring incumbents find new patches, with the initial quality level determined by the discovering firm's prior patch scaled by Λ. The third subtracts product obsolescence. The fourth is the net contribution of firm turnover: at rate δ_F a firm dies and an entrant takes its slot. Because the entrant brings one small product (weight $\bar{\Lambda}_E^{\sigma-1}$) in place of a whole firm (weight $\chi_{\mathcal{N}}$), firm turnover is a drag on average-quality growth.

Labor market clearing. R&D labor demand is proportional to quality weight. Using the firm shares and the product-to-firm ratio, aggregate labor used by exploiting and exploring firms is

$$L_I = \frac{c_I}{\sigma-1}\frac{f_I}{\chi_{\mathcal{N}}}\bar{z}_I, \qquad L_E = \frac{c_E}{\sigma-1}\frac{f_E}{\chi_{\mathcal{N}}}\bar{z}_E. \quad (30)$$

Entry uses $K_R E = K_R\delta_F F$ units of labor. Labor-market clearing is therefore $L = L^P + L_I + L_E + K_R\delta_F F$, or

$$F = \frac{L - L^P - L_I - L_E}{K_R\delta_F}. \quad (31)$$

Free entry. An entrant pays K_R units of labor, so the entry cost is $K_R w$. It begins with one product of quality $q_0 = \bar{\Lambda}_E\bar{Q}$. Let $\bar{W}_{\text{entry}} \equiv \sum_{k=1}^{K} p_k\hat{W}(x_{0,k})$ be the normalized value of starting exploitation on a newly drawn patch. Since $(q_0)^{\sigma-1} = \bar{\Lambda}_E^{\sigma-1}\bar{Q}^{\sigma-1} = \bar{\Lambda}_E^{\sigma-1}\frac{Q^{\sigma-1}}{\mathcal{N}}$, and $\kappa Q^{\sigma-1}/w = L^P/(\sigma-1)$, the free-entry condition (which holds with equality, since entry is strictly positive on the BGP: $E = \delta_F F > 0$) is $\frac{L^P}{\sigma-1}\frac{\bar{\Lambda}_E^{\sigma-1}}{\mathcal{N}}\bar{W}_{\text{entry}} = K_R$. Substituting $\mathcal{N} = \chi_{\mathcal{N}}F$ and using (31) gives the free-entry condition

$$\chi_{\mathcal{N}}\left[L - L^P - L_I - L_E\right] = \frac{L^P\delta_F}{\sigma-1}\bar{\Lambda}_E^{\sigma-1}\bar{W}_{\text{entry}}. \quad (32)$$

This is the second condition in the inner equilibrium block.

Definition 1 (Balanced growth path equilibrium)**.** *A balanced-growth-path equilibrium is a stationary allocation in normalized variables. It consists of a growth rate g, an interest rate r, production labor L^P, firm mass F, normalized firm values and policy $(\hat{W}, \hat{w}^E, x^*)$, and an average-quality path $\{\bar{Q}_t\}_{t\geq 0}$ with $\bar{Q}_t = \bar{Q}_0 e^{gt}$ for some $\bar{Q}_0 > 0$. It also includes the stationary composition and quality-weight objects induced by the firm policy, including $(F^I, F^E, E, \mathcal{N}, \chi_{\mathcal{N}})$ and $(\bar{z}_{\text{entry}}, \bar{z}_I, \bar{z}_E)$. These objects satisfy the following conditions:*

1. *Aggregate paths are implied by the stationary product mass $\mathcal{N}$ and the average-quality path:*

$$Q_t = \mathcal{N}^{1/(\sigma-1)}\bar{Q}_t, \qquad w_t = \frac{\sigma - 1}{\sigma} Q_t, \qquad Y_t = Q_t L^P, \qquad C_t = Y_t.$$

   The household Euler equation gives $r = \rho + g$, and normalized firm values are discounted at $\hat{\rho} = \rho + (\sigma - 1)g$ as in (8).

2. *Given (g, L^P), firms solve (10)–(13). The optimal policy is summarized by the cutoff x^* and the normalized values $(\hat{W}, \hat{w}^E)$.*

3. *The stationary firm-composition and variety objects generated by this policy satisfy (19)–(20), entry replaces firm death according to (21), and product-variety flows are stationary, $\dot{\mathcal{N}} = 0$ in (22).*

4. *The induced average-quality-weight process is stationary and finite. The moments defined in Appendix B.1 imply $(\bar{z}_{\text{entry}}, \bar{z}_I, \bar{z}_E)$ through (24)–(26), and satisfy the contractivity condition (27).*

5. *The aggregate quality scale is consistent with the stationary distribution of average-quality weights. Equivalently, the CES normalization (28) holds.*

6. *Production, R&D, and entry use the aggregate labor endowment: firm mass satisfies the labor-market condition (31), and entry satisfies the free-entry condition (32).*

3.3 Analytical characterization

How long does a firm exploit a patch? What determines the rate of growth in this economy? The equilibrium system of the previous subsection has no closed form and must be solved numerically. To build intuition, we step back and use an approximation method to analytically characterize a single firm's foraging problem and the resulting growth rate in partial equilibrium.

Sufficient conditions for g to coincide with g_q. Our analytical approximation applies to the firm-level renewal-reward rate g_q: the average rate at which a firm's log quality grows along its own innovation cycle. Under a simple set of sufficient conditions the aggregate growth rate g coincides with the firm-level rate g_q.

Proposition 1. *In an economy with (i) $\sigma \to 1$ (the CES index reduces to a geometric mean), (ii) $\delta_F = 0$ (no firm death), and (iii) no legacy products (the quality index only contains varieties from active patches), the aggregate growth rate g is equal to the per-firm renewal-reward rate g_q.*

Proof. Appendix B.4 presents a formal proof. To express the result as a well-defined limit of primitive parameters, we consider a version of the model extended to feature a depreciation rate for legacy products, δ_L, and condition (iii) corresponds to $\delta_L \to \infty$. □

The proposition thus also clarifies three wedges between g and g_q in the more general case. Each of the three conditions eliminates one of them. First, for $\sigma > 1$ the growth rate of the CES power mean picks up the cross-sectional variance of log quality, not just its mean, and that variance term has no reason to equal g_q. Under $\sigma \to 1$, log quality aggregates linearly. Second, with $\delta_F > 0$, a firm that dies during exploration is replaced by an entrant whose quality is proportional to the current mean $\bar{Q}$. But the firm it replaces was searching with a "frozen" quality that has fallen behind the rising mean. Death-and-replacement-during-exploration therefore averts part of the quality dilution that searching firms impose on the index. Third, the renewal-reward formula tracks the quality growth of a firm's active innovation line, whereas the economy-wide average quality index runs over all products, including "legacy products." Hence, with $\delta_L < \infty$, the quality index carries abandoned products whose quality is frozen, i.e., terms not growing at g_q, so the index lags behind frontier growth. Under the three sufficient conditions, the wedges are eliminated and the closed-form characterization of firm-level g_q is directly informative about the *aggregate* growth rate g. More generally, we have a clear understanding of the three wedges that remain.

Approximation. We now solve for the optimal stopping rule and the implied growth rate in closed form. The key trick is to adopt a *myopic approximation*: we assume that innovation rents, defined as $\hat{\Omega}(X) \equiv \hat{W}(X) - \hat{W}_0$, are small relative to the base firm value, $\hat{\Omega} \ll \hat{W}_0$, so that the firm treats $\hat{W}$ as being approximately equal to the passive legacy value $\hat{W}_0$. This is an accurate approximation when the option value of exploiting a freshly found patch is small relative to base firm value. In addition, we suppose that depletion is deterministic ($\sigma_x = 0$); there is a single patch type, with initial density x_0; the step-size function is exponential, i.e., $\theta(X) = e^{\phi X}$; and there is no capability loss upon cluster switching ($\Lambda = 1$).

Three quick points on notation: First, values are in the normalized units of (10): a unit of economic quality yields flow gross profit 1 before R&D cost, so the profit shifter κ does not appear. Second, under the exponential form, the quality-step semi-elasticity is $\eta \equiv (\sigma-1)\phi$, so the economic-quality gain per patent is $\theta(X)^{\sigma-1} - 1 = e^{\eta X} - 1$. Third, the total exit rate during exploitation is $\delta_{\text{exit}} \equiv \delta + \delta_F$, and the discount-plus-hazard rate is $\hat{\rho}_\delta \equiv \hat{\rho} + \delta + \delta_F$, so the legacy value (9) reads $\hat{W}_0 = 1/\hat{\rho}_\delta$.

Firm policy. With these assumptions and notation defined, the following lemma characterizes the firm's optimal policy for when to abandon a patch.

Lemma 1. *Under the myopic approximation, with homogeneous patches, an exponential step-size function, and deterministic depletion:*
(a) The exploration value (11) *is*

$$\hat{w}^E = \frac{-\hat{c}_E + \lambda \hat{W}_0}{\hat{\rho} + \lambda + \delta_F}. \tag{33}$$

(b) The optimal stopping threshold x^* satisfies

$$u(x^*) = (\hat{\rho} + \delta_F)\, \hat{w}^E + \hat{c}_I, \tag{34}$$

where $u(X) \equiv \gamma \hat{W}_0[e^{\eta X} - 1] = \gamma \hat{W}_0[\theta(X)^{\sigma-1} - 1]$ is the flow innovation payoff.
(c) In closed form,

$$x^* = \frac{1}{\eta} \ln\left(1 + \frac{R_{\mathit{eff}}}{\gamma}\right), \tag{35}$$

where the effective outside-option rate is

$$R_{\mathit{eff}} = \underbrace{\frac{(\hat{\rho} + \delta_F)\lambda}{\hat{\rho} + \lambda + \delta_F}}_{\text{exploration option}} + \underbrace{\hat{\rho}_\delta\, \hat{c}_I}_{\text{R\&D cost}} - \underbrace{\frac{(\hat{\rho} + \delta_F)\,\hat{\rho}_\delta\, \hat{c}_E}{\hat{\rho} + \lambda + \delta_F}}_{\text{search cost}}, \tag{36}$$

which is interior provided $R_{\mathit{eff}} > 0$, as we assume throughout.

Proof. (a) Start from the exploration value (11), specialized to $K = 1$ and $\Lambda = 1$:

$$(\hat{\rho} + \lambda + \delta_F)\, \hat{w}^E = -\hat{c}_E + \lambda \hat{W}(x_0).$$

Under the myopic approximation, substitute $\hat{W}_0$ for $\hat{W}(x_0)$ on the right and divide by $\hat{\rho} + \lambda + \delta_F$.
(b) Under the myopic approximation, the exploitation HJB (10) linearizes to

$$\mu\, \hat{\Omega}' = \hat{\rho}_\delta\, \hat{\Omega} - u(X) - (\delta\, \hat{w}^E - \hat{c}_I), \qquad X > x^*.$$

Evaluating at $X = x^*$ using value matching and smooth pasting (12)–(13), $\hat{\Omega}'(x^*) = 0$ and

$\hat{\Omega}(x^*) = \hat{w}^E$, so

$$0 = \hat{\rho}_\delta \, \hat{w}^E - u(x^*) - (\delta \, \hat{w}^E - \hat{c}_I) = (\hat{\rho} + \delta_F) \, \hat{w}^E - u(x^*) + \hat{c}_I.$$

Rearranging gives (34). Note δ cancels: at the stopping point creative destruction is neutral, since losing the product drops the firm into exploration, which already has value $\hat{w}^E$.
(c) Substitute $u(x^*) = \gamma \, \hat{W}_0[e^{\eta x^*} - 1]$, $\hat{W}_0 = 1/\hat{\rho}_\delta$, and (33) into (34) and solve for x^*. □

Condition (34) says the firm leaves when the flow innovation payoff falls to the opportunity cost of staying. The closed form (35) then shows what moves the threshold. A higher R_{eff} raises it, so the firm abandons the patch sooner: faster search or greater impatience raises the first term, costly exploitation the second, while costly search lowers R_{eff} through the third. Notice that the threshold depends on neither the depletion speed $|\mu|$ nor the initial density x_0: where the firm stops is separate from how fast the patch depletes and how rich it starts.

Implied gain and cluster time. With the optimal stopping strategy in hand, we can characterize the firm's renewal-reward growth rate g_q:

$$g_q = \frac{E[H]}{E[T]}, \tag{37}$$

where $E[H]$ is the expected log-quality gain from patents during one exploitation spell and $E[T]$ is the expected time from one spell's start to the next.

Consider the denominator first. If exploitation runs until the optimal stopping time, and with deterministic depletion ($\sigma_x = 0$), the firm reaches the threshold x^* after time $T \equiv \Delta/|\mu|$, where $\Delta \equiv x_0 - x^*$ is the usable patch depth. But not every spell runs to completion; exit hits at total rate δ_{exit}, so the firm survives to time t with probability $e^{-\delta_{\text{exit}} t}$. Denote by $P \equiv e^{-\delta_{\text{exit}} T}$ the probability of surviving the full spell. The cycle has two phases. In expectation, the exploitation phase lasts $E[T_I] = (1 - P)/\delta_{\text{exit}}$, where $(1 - P)$ absorbs early exit. With probability $P_{\text{expl}} = 1 - \delta_F E[T_I]$ the firm reaches exploration (rather than dying during exploitation), then waits an expected $1/(\lambda + \delta_F)$ for the next patch. Hence

$$E[T] = \frac{1 - P}{\delta_{\text{exit}}} + \frac{P_{\text{expl}}}{\lambda + \delta_F}. \tag{38}$$

To find the expected exploitation gain, note that at each instant t during exploitation, patents arrive at rate γ and each adds $\phi X(t) = \phi(x_0 - |\mu| t)$ to log quality. Integrating against the survival probability $e^{-\delta_{\text{exit}} t}$ gives:

$$E[H] = \frac{\gamma\phi}{\delta_{\text{exit}}}\left[\left(x_0 + \frac{\mu}{\delta_{\text{exit}}}\right)(1-P) + \Delta \cdot P\right]. \tag{39}$$

The two bracket terms split the quality contribution by outcome—spells cut short by exit $(1-P)$ and completed spells (P).

To sharpen the result further, suppose that $\delta_{\text{exit}} = 0$. Then every spell runs to completion: $P = 1$, $E[T_I] = T$, $P_{\text{expl}} = 1$; mean cycle duration is $E[T] = T + 1/\lambda$, with mean cluster duration $T = \Delta/|\mu|$, and the firm's switching frequency is the inverse, $1/\bar{T} = |\mu|\lambda/(\lambda\Delta + |\mu|)$. The quality-gain formula (39) reduces (by L'Hôpital as $\delta_{\text{exit}} \to 0$) to the midpoint rule, $E[H] = \gamma\phi\, T \cdot \frac{x_0 + x^*}{2}$. Hence, the growth rate factors as

$$g_q = \underbrace{\gamma\phi\,\frac{x_0 + x^*}{2}}_{g_{\text{exploit}}} \cdot \underbrace{\frac{\lambda\Delta}{\lambda\Delta + |\mu|}}_{f_I}. \tag{40}$$

The first product term, g_{exploit}, is the *within-cluster productivity*: the average rate of log-quality improvement while the firm exploits a patch. It depends on the (endogenous) depth of exploitation, via the midpoint density $(x_0 + x^*)/2$, and the patent parameters γ, ϕ, but not on the depletion speed $|\mu|$. The second term, f_I, is the *exploitation time share*: the fraction of the cycle spent exploiting rather than searching.

Straightforward differentiation then reveals how each parameter moves g_q—the pace of growth—and the duration of spells—the mode of growth—in partial equilibrium:

- More frequent findings ($\gamma \uparrow$) and a higher step elasticity ($\phi \uparrow$) raise g_{exploit} and hence g_q, and make exploitation more valuable at any density, so the firm stays longer.

- Faster depletion ($|\mu| \uparrow$) leaves the stopping rule untouched—x^* compares flow payoffs, not speeds—so it works only through the time share: $dg_q/d|\mu| = -g_{\text{exploit}}\,\lambda\Delta/(\lambda\Delta + |\mu|)^2 < 0$. Duration shortens and growth falls.

- A higher patch-finding rate ($\lambda \uparrow$) raises the exploration value, making the firm pickier (x^* up, hence g_{exploit} up). Spells unambiguously shorten. Growth rises if $\lambda\Delta^2 > 2\,|\mu|\,x^*$, which holds when the firm exhausts most of the patch before leaving ($x^* \ll x_0$). In our numerical exercises, $\partial g_q/\partial\lambda > 0$ throughout.

In sum, growth can be low either because the firm produces few quality gains on a patch (g_{exploit} small) or because it loses a lot of time between patches (f_I small). In the next section, we turn from qualitative characterization to quantitative statements.

4 The composition and pace of growth

In this section, we first attach numbers to the structural parameters, then use the calibrated model to study two questions. The first concerns the *composition* of growth: how much of aggregate growth originates in technology clusters new to the firms exploiting them? The second concerns its *pace*: when growth slows—or, as many now hope in the age of artificial intelligence (AI), accelerates—does the change originate in the conditions for exploitation or for exploration?

4.1 Illustrative calibration

We propose a simple calibration of our model's parameters. The goal is to capture key magnitudes while retaining maximal transparency. Readers primarily interested in the results can skip this subsection and refer to Table 3 for an overview of the parametrization.

Model specification and normalizations. We consider a slightly simplified version of the model, designed to isolate the core mechanisms. First, we restrict the quality scaling parameter for cluster entry, Λ, to 1. Second, we set the cost parameters in both modes of search equal to one another (i.e., $c_I = c_E = c$). Together, these assumptions isolate the central trade-off between the immediate but diminishing rewards from exploitation and the time cost of exploration. Third, the diffusion of patch density, σ_x, is set to 0, so heterogeneity in the density of patches is driven solely by the ex-ante distribution of patch quality and the endogenous depth of foraging.

In terms of functional-form assumptions, we adopt a one-parameter polynomial form for the step size function, $\theta(X) = (1 + X)^{\phi}$, so a single parameter $\phi > 0$ governs how steeply the quality step rises with patch density X.[20] Further, we replace the non-parametric $(K, x_{0,k}, p_k)$ triplet with a continuous lognormal distribution, truncated to a bounded support.[21]

We normalize two objects. First, we set $L = 1$. Second, the state variable X is latent, with its scale not identified, so we normalize the mean of the initial patch densities to equal one. This normalization pins down the location parameter μ_ℓ and leaves the dispersion σ_ℓ as the distribution's only free parameter.

Externally calibrated. Two parameters are calibrated externally. The discount rate is set at $\rho = 0.05$, a conventional value, and the elasticity of substitution across products, σ, is set equal to 4, in line with Garcia-Macia *et al.* (2019) and Aghion *et al.* (2026).

[20] If the objective were to fit the data as well as possible, one could introduce a second parameter indexing $\theta(X)$ in order to separately discipline the intercept (and thus the step size when entering a typical patch) and the slope (and thereby the pace at which the step size diminishes with patch density).

[21] With the polynomial (unbounded) step-size function, an unbounded support for x_0 would make the exploration value and the quality-weight objects of Appendix B.1 infinite. All upper-tail boundary conditions are accordingly imposed at the support bound.

Offline estimated. Five parameters are estimated offline, four of them prior to the SMM procedure ($\gamma, \lambda, \delta, \delta_F$) and one of them ex post (K_R).

The patent arrival rate is set to match the (aggregate) ratio of the total number of patents over the total number of cluster-years, yielding $\gamma = 2.28$. Both exit rates are set to match evidence from the US Business Dynamics Statistics (BDS, 1981–2018): δ is the establishment exit rate at continuing firms, equal to 0.033, and δ_F the death rate of mature (age $16+$) firms, equal to 0.055.[22]

One of the most important parameters is the patch encounter rate λ. Conceptually, identification is straightforward. In extensive search the firm encounters a new patch at Poisson rate λ or dies at rate δ_F, so conditional on observing the firm reach a new cluster, the waiting time T_E is exponential with rate $\lambda + \delta_F$. Hence, $T_E \sim \text{Exp}(\lambda + \delta_F)$, and $\lambda = \frac{1}{\mathbb{E}[T_E]} - \delta_F$. The empirical requirement thus boils down to measuring the average wait time, conditional on observing the firm reach a new cluster, $E[T_E]$. The practical challenge is to measure the expected wait time between clusters given that firms are often active in multiple patches simultaneously. We isolate episodes in which a firm most plausibly operates as a single sequential line. Our baseline filter selects transitions where the firm holds one cluster, all prior engagements have ended, it goes fully dark—no patent in any cluster—and then enters a single new cluster. This is the closest empirical analogue to the model's idle search phase. We consider several alternative filters that progressively relax these restrictions; Appendix A.4.4 defines each filter and reports results across all of them. We construct wait times in the sample of all corporate assignees with 50+ patents, which provides sufficient power for clean single-line transitions. Results are similar when restricted to public firms. The baseline wait time is 4.9 years, which given the calibrated δ_F implies $\lambda \approx 0.15$.

Finally, entry cost K_R is backed out from the observed average firm size of 17 employees. In the model, $m_F = \frac{L}{F}$, so $F = L/m_F$ and $K_R = \frac{(L - L^P - L_I - L_E)m_F}{\delta_F L}$. Given $L = 1$, this becomes $K_R = \delta_F^{-1}(1 - L^P - L_I - L_E)m_F$.

Internally calibrated via SMM. The remaining five parameters, collected in the vector $\psi = [\phi,\ |\mu|,\ c,\ \sigma_\ell,\ \bar{\Lambda}_E]$, are jointly inferred by minimizing the criterion

$$\mathcal{G}(\psi) = \sum_{j=1}^{5} \left(\frac{\hat{m}_j - m_j(\psi)}{\frac{1}{2}|\hat{m}_j| + \frac{1}{2}|m_j(\psi)|} \right)^2 ,$$

[22]We condition on mature firms because the model features a size- and age-invariant hazard and we want to avoid applying to all firms a high hazard rate that, in the data, is driven by selection of bad firms into exit, which is a margin the model does not feature.

Table 3: Parameter values

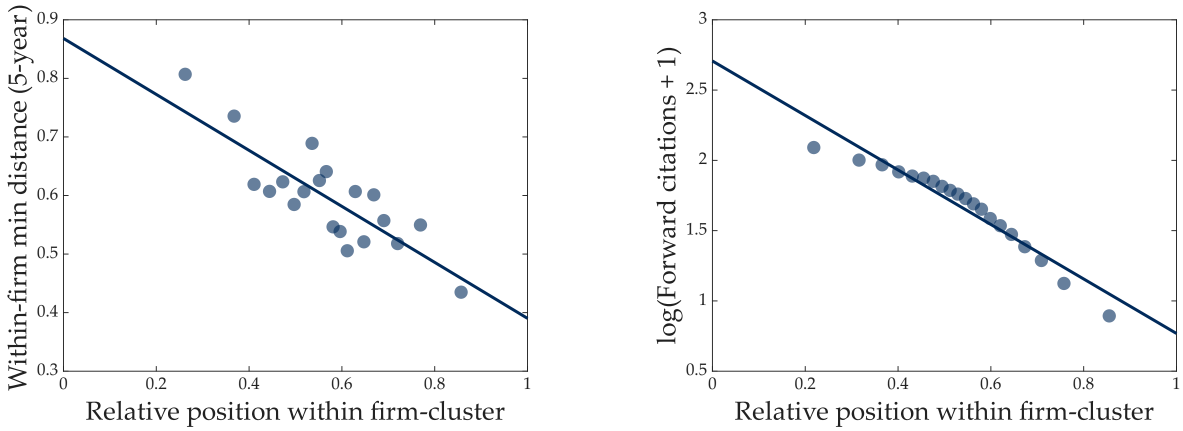
Parameter	Description	Value	Source / target	Data	Model
External					
ρ	Discount rate	0.05	Conventional value		
σ	Elasticity of substitution	4	Garcia-Macia et al. (2019)		
Offline					
γ	Patent arrival rate	2.28	Patents per cluster-year (patents)		
λ	Patch-finding rate	0.149	Inter-cluster wait (patents)		
δ	Product obsolescence rate	0.0331	Establishment exit (BDS)		
δ_F	Firm death rate	0.0548	Mature-firm death (BDS)		
K_R	Entry cost (labor)	22.0	Average firm size (BDS)		
Online/internal (SMM)					
ϕ	Patch quality slope	0.0321	Aggregate growth	0.020	0.020
$\lvert\mu\rvert$	Depletion speed	0.0681	Mean log active years (patents)	1.24	1.31
σ_ℓ	Initial-patch dispersion	0.353	Var. log patent count (patents)	1.17	1.17
c	R&D cost	0.177	R&D intensity (Compustat)	0.043	0.043
$\bar{\Lambda}_E$	Entrant relative quality	0.768	Entrant size ratio (BDS)	0.33	0.33

Notes. For the SMM-estimated group of parameters, each row lists the data moment most informative for that parameter alongside its model counterpart. Moments constructed from patent data use the baseline 50+ public firms dataset. The two cluster-shape moments use completed firm-cluster spells whose first and last patent both fall within 1981–2018. BDS moments are pooled over 1981–2018. R&D intensity is aggregate R&D expenditure divided by aggregate sales (xrd/sale) from Compustat, 1981–2018.

where $\hat{m}_j$ is the empirical moment and $m_j(\psi)$ its model counterpart. The five moments enter with equal weight, and the symmetric denominator places each percentage deviation on a comparable scale.

While the elements of ψ are jointly estimated, each is closely informed by one of the targeted moments. First, the depletion parameter μ directly governs the mean log cluster duration, which in the data is 1.24 (about 3.5 years at the geometric mean). Second, the parameter ϕ indexing the step size function is set to match the aggregate growth rate, which we take to be roughly 2%. Third, σ_ℓ is informed by the dispersion of patent count across clusters, measured as the variance of log count over all clusters (including singletons). Fourth, we calibrate the R&D cost parameter c to match aggregate R&D intensity, specifically the ratio of R&D expenditure divided by sales in Compustat. Finally, entrant quality $\bar{\Lambda}_E$ is informed by the entrant-to-average employment ratio, which is 0.33 in the BDS.

Measurement layer parameters. A couple of other parameters govern the measurement of auxiliary objects, without having any influence on the equilibrium. For citations, we follow Akcigit and Kerr (2018) and assume a patent is cited by each later patent in the same cluster, independently, with probability proportional to its quality step $s = \theta(X)^{\sigma-1} - 1$, so more



(a) Minimum distance

(b) Forward citations

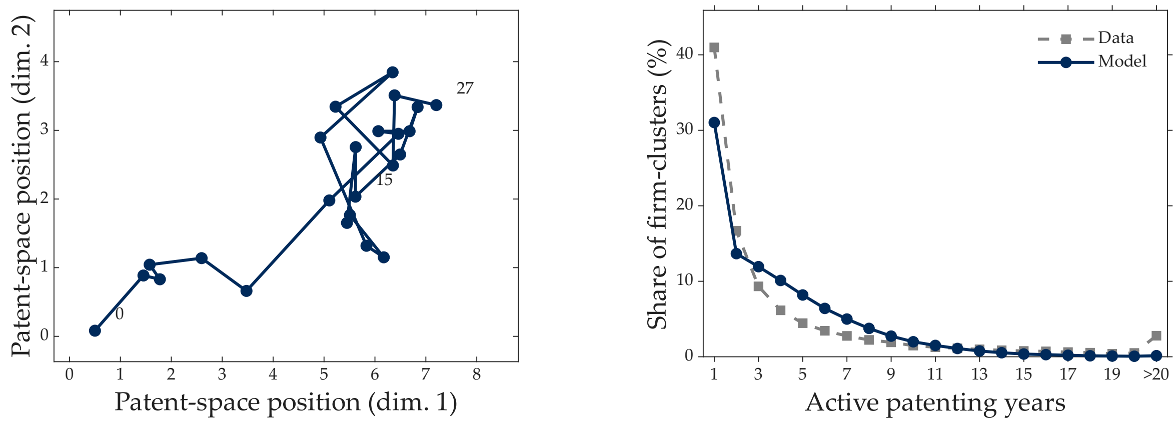
Figure 6: Diminishing returns within model clusters

Notes. Panel (a) plots the minimum distance to prior patents over relative position within a firm-cluster using simulated data. The minimum distance is measured within the firm's patent history. Panel (b) plots log forward citations. Both panels use firm-cluster and year fixed effects.

impactful (larger-step) patents accumulate more citations. We re-fit the single proportionality constant so the first patent in a cluster is cited with probability 0.75 on average, the value reported by Akcigit and Kerr (2018), and leave the rest of the citation distribution untargeted. For the spatial parameters—the within-patch diffusion σ_I and the exploration speed ν_E—we fix the unit from the data, measuring distances in the L2-normalized Euclidean patent-embedding metric. With $\alpha_d \equiv \frac{\lambda+\delta_F}{\nu_E}$, the mean distance traveled before entry satisfies $\bar{d}^E = 1/\alpha_d$, so $\nu_E = (\lambda+\delta_F)\bar{d}^E$. We measure $\bar{d}^E$ as the mean distance between the first patent in a newly entered cluster and the firm's immediately previous patent. For within-patch movement, between adjacent same-cluster patents, $E[r^2] = \frac{2\sigma_I^2}{\gamma}$, and $\sigma_I = \sqrt{\frac{1}{2}\gamma\,\overline{r^2}}$. We measure $\overline{r^2}$ as the mean squared distance between adjacent patents in the same firm-cluster sequence, ordered by application date and patent ID. The resulting moments imply $\nu_E = 0.234$ and $\sigma_I = 1.022$.

4.1.1 Parameter values and implied model properties

Table 3 summarizes the calibration. With five parameters and five targets the fit is almost exact, with a slight deviation on mean log active years that reflects a relative shortage of one-year cluster spells in the model. How can we interpret the values of key structural parameters? Given $\phi = 0.032$ in $\theta(X) = (1+X)^\phi$, a patent at the average fresh patch, with density normalized to one, raises product quality by $\theta(1) - 1 = 2.3\%$, and because θ rises in X these steps shrink as the patch depletes. With $|\mu| = 0.068$, depletion erodes about 6.8% of the average patch's density each year.



(a) Spatial path

(b) Active cluster durations

Figure 7: Spatial movement and cluster durations in the calibrated model

Notes. Panel (a) plots annual patent centroids for one simulated firm-life. Panel (b) compares model and data firm-cluster active durations. The model counts an entry-patent year for all simulated firm-clusters, including those with $x_0 < x^*$, so below-threshold draws have duration one.

To further gauge the implications of this calibration, we briefly revisit several of the empirical motivating facts presented in the introduction and Section 2 through the lens of the calibrated model. First, the model features diminishing returns at the patch level. In our calibration, we deliberately did not directly target the evidence on diminishing returns within clusters presented in Section 2.2.1—the only disciplining moment is cluster duration. Figure 6 reports the model counterpart to these moments, using minimum-distance as a proxy for novelty and citations as a proxy for quality. Both patent-level attributes decline in the relative position of the patent within the (firm) cluster—consistent with the motivating evidence for Fact 1.

Second, we look at the distribution of cluster duration. In the model, firms respond to heterogeneity in patch density by staying longer in richer patches, giving rise to a dispersed distribution of cluster spells. Panel (b) of Figure 7 compares the model-implied distribution of cluster duration with the empirical counterpart. While the model undershoots the data in the tails, i.e., clusters lasting only one year or ≥ 20 years, the overall shape is comparable.

Finally, to illustrate the spatial dynamics in the model, we draw a foraging path that echoes our opening discussion of Apple's trajectory in Figure 1. Panel (a) of Figure 7 depicts the path of a single foraging firm through the two-dimensional idea space.

Preview. In sum, under a very simple calibration, the model offers a plausible structural interpretation of the empirical facts. We next use it to study two applied questions, concerning different *modes* of growth and the determinants of the *pace* of growth.

4.2 Modes of growth: the vintage composition of innovation

In our model, the composition of innovation emerges as an equilibrium object. What, then, is the fraction of innovation—and ultimately of economic growth—that derives from firms working new patches rather than exploiting old ones?

Model-based answer. It is tempting to answer by pointing to the flow version of the aggregate growth equation (29), which at the calibrated parameters decomposes aggregate growth into four terms:

$$g = \underbrace{+0.0268}_{\text{quality improvement}} \underbrace{+0.0168}_{\text{discovery}} \underbrace{-0.0110}_{\text{obsolescence}} \underbrace{-0.0126}_{\text{net entry}} = 0.020.$$

Patenting on active products (g_{pat}) and the discovery of new patches are the two engines of quality creation; obsolescence and net firm entry are drags.[23] So is $+0.0168/0.02$ the share of net growth coming from new patches? This cannot be right, for two reasons. The discovery term reflects firm-specific capabilities accumulated through prior exploitation, so it does not cleanly map onto exploration. More important, because exploration initiates a new cycle of quality improvements, the quality-improvement term itself mixes patches of varying vintage—some old, others recent.

We therefore decompose patenting and growth by patch vintage, asking what fraction originates in patches the firm first entered within the past h years. Begin with the simplest version, which tracks how fast patenting rotates into fresh patches, ignoring differences in quality, or step size, across patents. Let $S_{\text{count}}(h)$ denote the fraction of patents over $(\tau, \tau + h]$ filed in patches first entered after the reference year τ. Appendix B.5 defines and characterizes this object formally. On the BGP it is stationary and depends only on cluster durations and the exit rate. Notably, $S_{\text{count}}(h) \to 1$ as $h \to \infty$: every cluster present at τ stops producing after a finite life, so the pre-existing cohort's cumulative contribution is bounded while new clusters keep arriving. Over a long enough horizon, *all* patenting comes from new areas—no patent filed in 2026 protects improvements to the design of DVD players.[24]

We find that a substantial share of patenting derives from areas the firm was not active in a decade or two earlier. We pool all patents, use the first-patent definition of cluster entry, and consider patents issued by all firms, whether incumbent at τ or born thereafter. Table 4 summarizes this and subsequent results. At the estimated parameters, $S_{\text{count}}(10) = 0.64$ and

[23] Net entry is negative because estimated entrant quality is low: an entrant brings a single product worth only 45.4% of average quality ($\bar{\Lambda}_E^{\sigma-1}$) in place of a whole incumbent firm, so each replacement lowers average quality.

[24] The comparative statics of S_{count} follow from the fact that shorter clusters raise the new-cluster share, together with the dependence of the stopping threshold x^* on the primitives: S_{count} is increasing in $|\mu|$ (patches deplete faster) and λ (exploration is more promising, so firms leave earlier), and decreasing in γ or ϕ (patents are more frequent or more valuable, so firms tolerate more depletion).

$S_{\text{count}}(20) = 0.81$.

The count share ignores patent quality, so it does not map to the growth rate. To connect with growth, and specifically with the decomposition in (29), we weight each patent by the associated incremental quality improvement $\Delta\tilde{q}$.[25] Summing these increments yields the quality-improvement component of growth g_{pat}:

$$(\sigma - 1)\, g_{\text{pat}} \;\equiv\; \frac{1}{\mathcal{N}} \int \gamma[\theta(x)^{\sigma-1} - 1]\, \Phi_z(x)\, dx \;=\; \frac{\Phi}{\tilde{Q}}, \qquad \Phi \equiv \sum_{\text{patents}} \Delta\tilde{q}, \;\; \tilde{q} \equiv q^{\sigma-1}, \tag{42}$$

where $\tilde{Q} \equiv Q^{\sigma-1} = \int_0^{\mathcal{N}} \tilde{q}_j \, dj$ is the aggregate economic-quality stock. The patenting flow Φ contributes $\Phi/\tilde{Q}$ to $(\sigma-1)g = \dot{\tilde{Q}}/\tilde{Q}$.[26] Thus, $S_{\text{qual}}(h)$ decomposes g_{pat} by cluster vintage.

Implementing this approach, Table 4 shows that, at a twenty-year horizon, patenting in new clusters accounts for almost two-thirds of quality improvement. The new-cluster quality share is $S_{\text{qual}}(10) = 0.44$ and $S_{\text{qual}}(20) = 0.65$. This lies below the count share because old clusters carry the firm's accumulated quality base; new-cluster patents take larger steps but on a smaller base.[27] What do these numbers imply for the origins of *net* growth g? Because $g_{\text{pat}} = 0.0268$ exceeds $g = 0.020$—discovery adds less than obsolescence and net entry subtract—new-cluster patenting translates, via the ratio $g_{\text{pat}}/g = 1.34$, into 88% of net growth at a twenty-year horizon. In short, sustained growth depends substantially on firms continually entering territory new to them.

An empirical sufficient statistic. If, per the preceding paragraphs, we can construct a growth composition by adding up patents of different vintages, can we perform this exercise directly in the data, without the full model structure and calibration? The primary obstacle is that the quality increment $\Delta\tilde{q}$ is unobserved, so S_{qual} is not directly measurable. The model resolves this by identifying an observable approximate sufficient statistic: the new-cluster share weighted by patent *market value*, as measured by Kogan *et al.* (2017). In the model, a patent raises

[25] More generally, weighting patents by w_p, the new-cluster share over $(\tau, \tau+h]$ is

$$S_{\text{new}}(h; w) = \frac{\sum_{p:\, y_p \in (\tau, \tau+h]} w_p \;\cdot\; \mathbf{1}[e(f_p, c_p) \geq \tau]}{\sum_{p:\, y_p \in (\tau, \tau+h]} w_p}. \tag{41}$$

[26] g_{pat} is a rate, so the common trend $e^{(\sigma-1)gt}$ cancels between the new-cluster flow and the total; $S_{\text{qual}}(h)$ averages the resulting stationary share over the window. Summing nominal increments would instead decompose the cumulative quality created, a level, and overstate the new share.

[27] Crediting the discovery term to new patches as well raises the share further. Including g_{disc} in the calculation transforms the quality share as

$$S_{\text{qual}}(h) \;\mapsto\; \frac{S_{\text{qual}}(h) + g_{\text{disc}}/g_{\text{pat}}}{1 + g_{\text{disc}}/g_{\text{pat}}}. \tag{43}$$

With $g_{\text{disc}}/g_{\text{pat}} = 0.63$, this lifts the twenty-year share from 0.65 to 0.79.

Table 4: Patch vintage composition of innovation and growth

	Model		Data	
Patent weight	$h = 10$	$h = 20$	$h = 10$	$h = 20$
A. Share of patenting from new clusters				
Count	0.64	0.81	0.37	0.51
Quality increment	0.44	0.65	–	–
Private value	0.45	0.66	0.40	0.49
B. Seed-inclusive and net-growth shares (model only)				
Quality + discovery seed	0.66	0.79		
New-cluster share of net growth	0.60	0.88		

Notes. This table shows the new-cluster share $S(h)$ over the window $(\tau, \tau + h]$ under alternative weighting schemes. Calculations in the data column use an average over multiple starting points, $\tau \in \{1985, \ldots, 1998\}$, and construct entry indicators using the peak-flow definition. Throughout, we pool patents across firms. A dash marks a weight with no counterpart on that side. Panel B is model-only: the seed-inclusive share adds the discovery seed via equation (43), and the net-growth share is $(g_{\text{pat}}/g)\, S_{\text{qual}}(h)$, with $g_{\text{pat}} = 0.0268$, $g = 0.020$, $g_{\text{pat}}/g = 1.34$.

firm value by $\Delta V = \kappa\, \Delta\tilde{q}\, \hat{W}(X)$, where $\hat{W}(X)$ is the normalized value of the active line plus continuation (equation (10))—precisely the patent-triggered revaluation that KPSS measure. Up to the common time-specific factor κ, the wedge between $\Delta\tilde{q}$ and ΔV is $\hat{W}(X)$, which varies little across X: a fresher patch carries more option value, but that premium is small relative to the X-independent value of the quality already in hand. In the calibrated model, S_{val} therefore tracks S_{qual} to within a percentage point (Table 4).

Empirical implementation raises two challenges absent from the model. First, a single patent assigned to a cluster in a given year suffices to mark it as "new," even if the assignment reflects noise or follow-up patenting arrives only a decade later. We therefore date cluster entry by the peak-flow rule rather than the first-patent event. Second, the data exhibit non-stationarities—KPSS values spike around the dot-com bubble, unlikely to reflect a genuine rise in quality improvements—so we average across reference dates $\tau \in \{1985, \ldots, 1998\}$.[28]

Table 4 reports the result. According to the patent-value proxy, new clusters account for about half of quality improvement growth over a twenty-year horizon.[29] It sits below the model-based share, but the two agree in order of magnitude.

[28]Manual inspection suggests the peak-flow rule still dates "cluster birth" well before the bulk of the patent flow, which biases the measured share upward at short horizons. Over long horizons, a few large incumbents ride old clusters far longer than the model's durations allow, so matching the model's new-cluster share becomes harder as h grows; weighting firms equally rather than pooling raises the empirical share well above the pooled version.

[29]The patent-value-based share is probably best read as the data counterpart to the vintage decomposition of the patenting-increment flow g_{pat}. KPSS prices the revaluation at a patent's grant—the increment that patent embodies—whereas the seed, the quality a firm carries into a new area, is not itself a patent. Insofar as entry is marked by a patent that also prices this carried-in quality, the KPSS share corresponds to the seed-inclusive object (43) rather than to g_{pat} alone.

Table 5: Modes of growth

	$\lvert\mu\rvert$	λ	g	$S_{\text{qual}}(10)$	$S_{\text{qual}}(20)$	Avg. dur.	g_{pat}/g	g_{disc}/g	g/L_R	$L_E \mid L_I$
Baseline	0.068	0.149	2.0%	0.44	0.65	3.72	134%	84%	0.397	0.020 | 0.030
Counterfact.	0.057	0.068	2.0%	0.34↓	0.55	4.63	161%	44%	0.318	0.023 | 0.040

Notes. This table compares the baseline economy to a counterfactual one that attains the same growth rate g but with a 10ppt lower share of g_{pat} derived from new clusters at horizon $h = 10$, S_{qual}; the ↓ in the corresponding column marks this targeted reduction. The main text describes how this counterfactual is constructed. $S_{\text{qual}}(h)$ is the stationary $\tilde{q}$ patenting quality-improvement flow share from patches discovered in the last h years; Avg. dur. is $\exp(E[\log(\text{active years})])$ cond. on patches having ≥ 2 patents; g_{pat}/g and g_{disc}/g are the patenting and discovery contributions to aggregate growth g (net entry and obsolescence, both negative offsets, are omitted, so the two shown exceed 100%).

Modes of growth. How should we interpret the vintage composition? If we compare our baseline economy, calibrated to U.S. data, with a sibling economy with a lower new-cluster share, is the latter less dynamic? The answer is subtle. A lower new-cluster share unambiguously connotes slower technological turnover, but not necessarily more feeble growth. It may reflect difficult exploration, where firms deepen depleted patches for lack of better options. But it may equally reflect effective exploitation, where patches deplete slowly and firms rationally concentrate on rich existing territory. The two economies can sustain the same aggregate growth rate through different *modes* of growth.

To make this concrete, we construct a counterfactual economy by lowering both the patch-finding rate λ and the within-patch depletion rate $\lvert\mu\rvert$, choosing the point along a fixed-g isoquant such that $S_{\text{qual}}(10)$ falls by 10 percentage points relative to baseline. The two changes move composition in the same direction but growth in opposite directions. A lower $\lvert\mu\rvert$ raises growth and directly lowers the new-cluster share, as firms exploit locally for longer. A lower λ lowers growth and likewise extends cluster duration, as firms respond to the worse exploration payoff by exploiting longer. Along the isoquant, the two growth effects cancel. In the counterfactual, $\lvert\mu\rvert$ is 16% lower and λ is 54% lower: patches deplete more slowly, but exploration is less promising.

Table 5 compares the two economies. In the "higher-exploitation" economy, firms typically stay active in any one cluster for longer. Achieving the same growth with less exploration requires more R&D labor, so research productivity—growth per unit of R&D—falls. Both types of R&D labor rise, especially for exploitation, as firms reallocate toward the now relatively more productive mode. Decomposing net growth, discovery contributes less, offset by higher quality improvements. Firms explore less new territory but squeeze more out of the patches they hold.

In closing, it bears noting that this counterfactual exercise echoes the varieties-of-capitalism debate in political economy (Hall and Soskice, 2001).[30] "Liberal market economies" would

[30] Hall and Soskice (2001) argue that "coordinated market economies" have a comparative advantage in incremental innovation, marked by continuous small-scale improvements to existing product lines and production processes,

correspond to higher λ and $|\mu|$; "coordinated market economies" to lower values. This connection raises questions an extended version of our framework could take up. Are some financial market arrangements more conducive to growth when λ and $|\mu|$ are high? Are labor market institutions that provide employment protection more beneficial when growth relies on exploitation?

To recap, this section established two results. First, at a twenty-year horizon, patches new to the firm account for roughly two-thirds of quality-improvement growth in the model—about half in the value-weighted data—and an even larger share of net growth. Sustained growth requires firms continually breaking new ground. Second, two economies can grow at the same rate while differing sharply in how much of that growth comes from new territory.

4.3 The pace of growth: slowdowns and accelerations

The previous subsection held aggregate growth fixed and asked how its composition can differ. Here we ask what moves the pace of growth itself. At the core of the analysis stands a sharp model-implied diagnostic: the duration of patch spells and the rate at which firms enter new patches respond in opposite directions to worsening exploitation and worsening exploration, even though both lower growth. We apply this diagnostic as an identification tool in two contexts. Looking back: did growth slow because firms' *exploitation* of the technologies they hold deteriorated, or because *exploration* for new ones became harder? Looking ahead: what kind of "method of invention" is AI?

Growth slowdowns. Measured productivity growth in the U.S. and other advanced economies has slowed for several decades, and a large literature argues that impactful innovation has become harder: ideas are getting harder to find (Bloom *et al.*, 2020), patents are less disruptive (Park *et al.*, 2023), the burden of knowledge is rising (Jones, 2009). The identification problem is that the pace alone is silent about the cause: every deterioration the literature has proposed lowers the growth rate and measured research productivity alike.

Through the lens of our model—recall the analytical results from Section 3.3—a growth *slowdown* can come from either side of the firm's problem. Exploitation can deteriorate: patches deplete faster ($|\mu| \uparrow$), each patent delivers a smaller quality step ($\phi \downarrow$), or patents arrive less often ($\gamma \downarrow$). Or exploration can deteriorate: new patches become harder to find ($\lambda \downarrow$). Each lowers growth.

The sign of changes in the duration of cluster spells, however, can tell these forces apart. When exploitation worsens, the firm reaches its quitting threshold sooner, or sits on a patch that runs down faster: spells shorten, and the firm moves across clusters more rapidly. When

while "liberal market economies" are more supportive of radical innovation, involving substantial shifts in product lines or entirely new goods.

Table 6: Structural drivers of the pace of growth: empirical signatures

	g (%)	g/R	g_q/R	R&D int. (%)	duration (yr)	entry rate	patents/L_I
Baseline	1.99	0.40	0.50	4.3	3.94	0.128	2.24
Exploration worsens							
$\lambda \downarrow 20\%$	1.89	0.35	0.45	4.6	4.40	0.114	2.22
Exploitation worsens							
$\lvert\mu\rvert \uparrow 20\%$	1.73	0.36	0.49	4.1	3.70	0.131	2.45
$\phi \downarrow 20\%$	1.18	0.26	0.42	3.9	3.52	0.134	2.54
$\gamma \downarrow 20\%$	1.20	0.26	0.42	3.9	3.52	0.134	2.02
Obsolescence rises							
$\delta \uparrow 20\%$	1.97	0.38	0.47	4.5	3.88	0.129	2.21

Notes. Each row lowers g by a 20% move in one primitive, with general equilibrium re-solved at the estimated parameters. Duration is mean firm-cluster exploitation length $\bar{\tau}_I$; the entry rate is new clusters entered per firm-year. The δ move is shown at 20% for comparability.

exploration worsens, the outside option falls, the firm lowers its threshold x^* and exploits each patch longer, and it enters fewer new clusters: spells lengthen and rotation across clusters slows.

Table 6 verifies and quantifies these predictions in the estimated general-equilibrium model. Each row lowers growth by moving one primitive by twenty percent. Research productivity g/R (and g_q/R) likewise falls in every row. So "ideas are getting harder to find," taken as the productivity ratio alone, is consistent with all the explanations. In contrast, *only* a fall in λ lengthens duration while lowering the entry rate, whereas every exploitation channel does the reverse.[31]

What, then, do the data tell us about the drivers of a growth slowdown? We measure duration as the active patenting years a firm records in a cluster, as before, but over a fixed window of ten years after entry to avoid unequal truncation over time. Table 7 shows no sustained downward trend in duration in the public sample; Table A.6 in the Appendix shows the same pattern in the wider set of all corporate patenters.[32] The rate at which firms enter clusters new to them has likely also fallen.[33]

[31]The patent rate patents/L_I adds a second contrast that differentiates between the exploitation channels themselves: it falls when patents grow scarcer (γ down) or lines die faster (δ up), but holds up or rises when faster depletion or a falling step size is the culprit.

[32]An alternative measure, the first-to-last calendar span, declines over the sample, but that decline is largely mechanical: early cohorts have more room to leave a cluster dormant and later return. Removing such gaps removes most of the shortening. The fixed-window measure avoids this right-censoring bias and is consistent with how we calibrated the model.

[33]We view the evidence as suggestive. A firm can enter a cluster for the first time only once, so the raw entry rate falls mechanically as firms accumulate clusters over their lifetimes. Holding firm age fixed—comparing firms in their first five patenting years across birth cohorts—the trend is roughly flat. Accounting for re-entry after a gap, which prevents portfolio history from accumulating, likewise attenuates the measured fall. See Appendix A.4.5 for details.

Table 7: Cluster duration and entry over time

Panel A. Conditional firm-cluster duration					
Object	Definition	1981–1989	1990–1999	2000–2009	2010–2015
Duration	First-patent date	3.16	3.97	3.68	3.56
Duration	Peak-flow date	3.40	4.17	4.01	3.86
Panel B. Entry and re-entry rates					
Object	Definition	1981–1989	1990–1999	2000–2009	2010–2018
First-ever entry rate	First-patent entry	1.43	1.32	1.10	0.75
Entry/re-entry rate	First entry or >5-year return	2.13	2.03	1.81	1.73
Panel C. Entry rates by firm birth cohort					
Object	Definition	1982–1989	1990–1999	2000–2009	
First-ever entry rate	First 5 patenting years	1.44	1.79	1.47	

Notes. All rows use the baseline sample and require at least two patents in the firm-cluster over its observed history. Panel A reports the mean number of patenting years within the ten-year window from entry, dated by the first patent in the cluster or by the peak-flow rule (footnote 12); columns are entry cohorts, the last ending in 2015 so every episode has a full window. Panel B reports events per active firm-year (a year with at least one clustered patent) by calendar year; re-entry adds returns after more than five years without patenting in that cluster. Panel C columns are firm birth cohorts, dated by the first clustered patent, restricting to firm-years within the firm's first five patenting years. See Appendix A.4.5 for details.

A duration that does not shorten is hard to square with a reading of the slowdown as worsening exploitation: faster depletion or smaller steps would have cut it. The R&D evidence points the same way: a decline in the return to exploitation ($\phi \downarrow$ or $\gamma \downarrow$) would lead to falling R&D intensity, whereas research effort has in fact risen (Bloom *et al.*, 2020).

In contrast, the evidence leans toward a deterioration in the exploration margin. We view the case as suggestive, given the measurement difficulties discussed above. That said, independent evidence points in the same direction. Kalyani (2024), whose text-based creativity measure we used earlier, documents a shift from creative toward derivative patents during 1950–2015. Going back further, Youn *et al.* (2015) find that the rate at which genuinely new technologies appear has fallen since the nineteenth century, even as invention has continued by recombining existing ones. In sum, the evidence weighs against a slowdown driven by deteriorating exploitation and is tentatively consistent with harder exploration.

Growth accelerations: what kind of "method of invention" is AI? If the past decades raised the fear that growth is slowing, recent advances in artificial intelligence raise the opposite hope. The theoretical case for an acceleration rests on a specific premise: that AI is not merely automation in the production of goods and services but a *method of invention*, a technology that helps produce ideas themselves (Cockburn *et al.*, 2018; Aghion *et al.*, 2017; Agrawal *et al.*, 2018; Jones, 2022). Taking this premise as given, however, views diverge on what it implies. Does AI "merely" accelerate search within domains researchers already work in—as, arguably, earlier technologies did (Aldighieri and Malpassi, 2025)? Or can AI's breadth of knowledge, beyond any single researcher's reach, open genuinely new territory? What evidence there is from early studies remains inconclusive.[34]

Our model makes the distinction between different conceptions of AI-in-innovation testable by applying the same identification logic prospectively. Suppose first that AI accelerates exploitation, raising the rate γ at which a firm obtains patents on patches it already holds. Then each patch is worth more, so spells lengthen, and the new-cluster share falls. Suppose, instead, that AI aids exploration, making new patches easier to find ($\lambda \uparrow$). This leads the firm to raise its threshold x^* and leave sooner, so spells shorten, and the new-cluster share rises. Notably, *both* channels typically raise g. An AI that accelerates exploitation thus need not be a lesser force. But it does leave growth hinging, as before, on humans continuing to find new patches. The distinction bears primarily on the composition of growth. If, as the evidence above suggests, the past slowdown reflects deteriorating exploration conditions, then an AI confined to exploitation would restore the growth rate while narrowing innovation still further to familiar territory.

5 Concluding discussion: limitations and future research

Foraging animals traverse space in a distinctive rhythm: many small steps within a gradually depleting patch, followed by an occasional long flight to the next. In this paper, we applied the logic of optimal foraging to firms' search for valuable ideas. In four decades of patent text, we found evidence of the same rhythm—diminishing returns within patches of related ideas, long spells of local refinement, occasional jumps that renew a firm's innovation—and built a theory of endogenous growth around it.

Our aim was to develop a transparent foundation for a foraging theory of growth: a framework that isolates the explore–exploit margin, allows for closed-form characterizations that clarify mechanisms, and speaks to data. That transparency rests on simplifying assumptions. We close by making four of them explicit, both to delimit what our results establish and to mark promising lines of further research.

First, firms in our model interact only through prices and free entry. There is no explicit crowding or business stealing at the patch level; the exogenous depletion drift and destruction hazard absorb such forces in reduced form. One consequence is that every patch a firm enters is

[34] In chemistry, Boiko *et al.* (2023) find that autonomous laboratories speed up experimentation on known compound classes by orders of magnitude. AlphaFold uses deep learning to predict protein structures, accomplishing in seconds what previously required years of experimental crystallography. Hill and Stein (2026) show that research on previously unsolved proteins rose 15–40% after its release, suggesting AI can redirect scientific attention toward territory that was effectively inaccessible before (though no downstream effect on drug development has appeared so far). Using firm-level data, Babina *et al.* (2024) find evidence that AI adopters expand into new product areas.

new to the economy, so the firm-level entry margin coincides with the aggregate expansion of the idea space. A richer industrial organization of the patch is a natural next step. When several firms can occupy the same patch, do they race for its best ideas and deplete it for one another? Making depletion and destruction rates equilibrium objects would yield a fuller account of the composition and pace of growth.

Second, and relatedly, we abstracted from knowledge spillovers. Section 2 showed that a patch's earliest patents are disproportionately likely to be highly cited. Our model reads these citations as a signal of quality, without separating private from social value. A long tradition instead reads citations as the paper trail of knowledge spillovers (Jaffe *et al.*, 1993; Bloom *et al.*, 2013). On that reading, our evidence suggests early-in-patch patents carry positive externalities. A natural extension of our model would make exploration a form of pioneering: the discovery of a patch new to the economy raises the rate at which other explorers encounter new patches. In such an extension, exploration is plausibly undersupplied in equilibrium, pointing to a case for subsidizing exploration specifically rather than R&D generally.

Third, in the model, each firm improves a single active line. In reality, large firms improve several at once. The simplest extension of our model would allow firms to have multiple independent slots. It would also be interesting, though technically demanding, to allow for interdependencies, whereby success on one patch draws resources and attention away from others. Contemporary AI research offers an example: the rise of large language models has concentrated researchers and capital on one patch, while adjacent directions are neglected.

Fourth, our innovating agent is the firm, but the theory extends naturally to individual scientists and inventors. The empirical literature suggests that papers and patents become less disruptive as their authors age (Kaltenberg *et al.*, 2023; Cui *et al.*, 2026). While this may reflect innate ability declining with age, our foraging model points to an economic mechanism: exploration is an investment whose payoff accrues over the remaining career, so a shortening horizon lowers the incentive to explore. Distinguishing between these accounts—ability versus incentives—is a first-order question for the economics of science and R&D policy.

References

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, Reallocation, and Growth. *American Economic Review*, **108**(11), 3450–3491.

Acemoglu, D., Akcigit, U., and Celik, M. A. (2022). Radical and Incremental Innovation: The Roles of Firms, Managers, and Innovators. *American Economic Journal: Macroeconomics*, **14**(3), 199–249.

Aghion, P. and Howitt, P. (1992). A Model of Growth through Creative Destruction. *Econometrica*, **60**(2), 323–51.

Aghion, P., Jones, B. F., and Jones, C. I. (2017). Artificial Intelligence and Economic Growth.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2023). A Theory of Falling Growth and Rising Rents. *The Review of Economic Studies*, **90**(6), 2675–2702.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2025). Good Rents versus Bad Rents: R&D Misallocation and Growth.

Aghion, P., Bergeaud, A., Boppart, T., and Brouillette, J.-F. (2026). Resetting the Innovation Clock: Endogenous Growth through Technological Turnover. *American Economic Review: Insights*.

Agrawal, A., McHale, J., and Oettl, A. (2018). Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth.

Akcigit, U. and Kerr, W. R. (2018). Growth through Heterogeneous Innovations. *Journal of Political Economy*, **126**(4), 1374–1443.

Akcigit, U., Hanley, D., and Serrano-Velarde, N. (2021). Back to Basics: Basic Research Spillovers, Innovation Policy, and Growth. *The Review of Economic Studies*, **88**(1), 1–43.

Aldighieri, P. and Malpassi, F. (2025). Technological Breakthroughs and the Progress of Science: Evidence from Early Computers.

Argente, D., Lee, M., and Moreira, S. (2024). The Life Cycle of Products: Evidence and Implications. *Journal of Political Economy*, **132**(2), 337–390.

Babina, T., Fedyk, A., He, A., and Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, **151**, 103745.

Bénichou, O. (2006). Two-dimensional intermittent search processes: An alternative to Lévy flight strategies. *Physical Review E*, **74**(2).

Bénichou, O., Coppey, M., Moreau, M., Suet, P.-H., and Voituriez, R. (2005). Optimal Search Strategies for Hidden Targets. *Physical Review Letters*, **94**(19), 198101.

Bénichou, O., Loverdo, C., Moreau, M., and Voituriez, R. (2011). Intermittent search strategies. *Reviews of Modern Physics*, **83**(1), 81–129.

Berlingieri, G., Ridder, M. D., Lashkari, D., and Rigo, D. (2025). Creative destruction through innovation bursts. *CEP Discussion Papers*, (dp2095).

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying Technology Spillovers and Product Market Rivalry. *Econometrica*, **81**(4), 1347–1393.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are Ideas Getting Harder to Find? *American Economic Review*, **110**(4), 1104–1144.

Boiko, D. A., MacKnight, R., Kline, B., and Gomes, G. (2023). Autonomous chemical research with large language models. *Nature*, **624**(7992), 570–578.

Buera, F. J. and Oberfield, E. (2020). The Global Diffusion of Ideas. *Econometrica*, **88**(1), 83–114.

Callander, S. (2011). Searching and Learning by Trial and Error. *American Economic Review*, **101**(6), 2277–2308.

Carnehl, C. and Schneider, J. (2025). A Quest for Knowledge. *Econometrica*, **93**(2), 623–659.

Carvalho, V., Draca, M., and Kuhlen, N. (2021). Exploration and Exploitation in US Technological Change. Working paper, Competitive Advantage in the Global Economy (CAGE).

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, **9**(2), 129–136.

Choi, J., Braguinsky, S., Ding, Y., Jo, K., and Kim, S. (2023). Mega Firms and New Technological Trajectories in the U.S.

Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press, Boston.

Cockburn, I. M., Henderson, R., and Stern, S. (2018). The Impact of Artificial Intelligence on Innovation.

Cowie, R. J. (1977). Optimal foraging in great tits (Parus major). *Nature*, **268**(5616), 137–139.

Cui, H., Lin, Y., Wu, L., and Evans, J. A. (2026). Aging and the narrowing of scientific innovation. *Science*, **392**(6798), 588–591.

Cyert, R. M. and March, J. G. (1963). *A Behavioral Theory of the Firm*. A Behavioral Theory of the Firm. Prentice Hall/Pearson Education, Upper Saddle River, NJ, US.

De Ridder, M. (2024). Market Power and Innovation in the Intangible Economy. *American Economic Review*, **114**(1), 199–251.

Fort, T. C., Goldschlag, N., Liang, J., Schott, P. K., and Zolas, N. (2026). Growth is Getting Harder to Find, Not Ideas.

Ganguli, I., Lin, J., Meursault, V., and Reynolds, N. (2026). Spreading Out Across Expanding Idea Space.

Garcia-Macia, D., Hsieh, C.-T., and Klenow, P. J. (2019). How Destructive Is Innovation? *Econometrica*, **87**(5), 1507–1541.

Grossman, G. M. and Helpman, E. (1991). Quality Ladders in the Theory of Growth. *The Review of Economic Studies*, **58**(1), 43–61.

Hall, P. A. and Soskice, D., editors (2001). *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. OUP Oxford, Oxford England ; New York.

Hill, R. and Stein, C. (2026). How Artificial Intelligence Shapes Science: Evidence from AlphaFold.

Hopenhayn, H. and Squintani, F. (2021). On the Direction of Innovation. *Journal of Political Economy*, **129**(7), 1991–2022.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, **108**(3), 577–598.

Jones, B. F. (2009). The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *The Review of Economic Studies*, **76**(1), 283–317.

Jones, C. I. (2022). The Past and Future of Economic Growth: A Semi-Endogenous Perspective. *Annual Review of Economics*, **14**(Volume 14, 2022), 125–152.

Jovanovic, B. and Rob, R. (1990). Long Waves and Short Waves: Growth Through Intensive and Extensive Search. *Econometrica*, **58**(6), 1391–1409.

Jovanovic, B. and Rousseau, P. L. (2001). Why Wait? A Century of Life before IPO. *American Economic Review*, **91**(2), 336–341.

Kaltenberg, M., Jaffe, A. B., and Lachman, M. E. (2023). Invention and the life course: Age differences in patenting. *Research Policy*, **52**(1), 104629.

Kalyani, A. (2024). The Creativity Decline: Evidence from US Patents.

Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M. (2021). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights*, **3**(3), 303–320.

Klette, T. J. and Kortum, S. (2004). Innovating Firms and Aggregate Innovation. *Journal of Political Economy*, **112**(5), 986–1018.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological Innovation, Resource Allocation, and Growth. *The Quarterly Journal of Economics*, **132**(2), 665–712.

Kortum, S. S. (1997). Research, Patenting, and Technological Change. *Econometrica*, **65**(6), 1389–1419.

Krebs, J. R., Ryan, J. C., and Charnov, E. L. (1974). Hunting by expectation or optimal foraging? A study of patch use by chickadees. *Animal Behaviour*, **22**, 953–IN3.

Lafond, F. and Kim, D. (2019). Long-run dynamics of the U.S. patent classification system. *Journal of Evolutionary Economics*, **29**(2), 631–664.

Loughran, T. and Ritter, J. (2004). Why Has IPO Underpricing Changed Over Time? *Financial Management*, **33**(3).

Lucas, R. E. and Moll, B. (2014). Knowledge Growth and the Allocation of Time. *Journal of Political Economy*, **122**(1), 1–51.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, **2**(1), 71–87.

Park, M., Leahey, E., and Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, **613**(7942), 138–144.

Perla, J. and Tonetti, C. (2014). Equilibrium Imitation and Growth. *Journal of Political Economy*, **122**(1), 52–76.

Peters, M. and Walsh, C. (2026). Population Growth and Firm Dynamics. *Journal of Political Economy Macroeconomics*, **4**(2), 256–296.

Stephens, D. W. and Krebs, J. R. (1986). *Foraging Theory*, volume 1. Princeton University Press.

Stokey, N. L. (2008). *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton University Press.

Weitzman, M. L. (1979). Optimal Search for the Best Alternative. *Econometrica*, **47**(3), 641–654.

Weitzman, M. L. (1998). Recombinant Growth. *The Quarterly Journal of Economics*, **113**(2), 331–360.

Youn, H., Strumsky, D., Bettencourt, L. M. A., and Lobo, J. (2015). Invention as a combinatorial process: Evidence from US patents. *Journal of The Royal Society Interface*, **12**(106), 20150272.

Online Appendix

This appendix contains supplemental material. Any references to sections, equations, figures, or tables that are not preceded by a capital letter refer to the main article.

Appendix A Appendix: Empirics

A.1 Data

This section describes the data sources, sample selection and variable construction. Appendix A.2 details the embedding and clustering procedures.

Patent data come from the PatentsView tables of granted U.S. patents. This source provides patent metadata, titles and abstracts, application and grant dates, disambiguated assignee identifiers, citation links, and CPC classifications. We merge in three external patent-level measures: real patent value from Kogan *et al.* (2017), (five-year) breakthrough importance from Kelly *et al.* (2021), and creativity from Kalyani (2024). Firm-level variables—company identifiers, sales, and R&D expenditures—come from the CRSP/Compustat merged database, linked to patents through the CRSP `permno` identifier supplied by Kogan *et al.* (2017). Firm IPO years combine CRSP/Compustat listing dates with Ritter's IPO database (Loughran and Ritter, 2004) and the historical listing data of Jovanovic and Rousseau (2001).

We keep granted utility patents with a valid application and grant date and a non-missing title and abstract, excluding withdrawn patents and patents listing multiple assignees, for application dates from 1900 through 2024 (coverage is material only from the 1970s onward). This yields 7,930,021 patents. Because the analysis treats each patent's text as describing a distinct idea, we then drop records whose combined title and abstract duplicates that of an earlier patent, exactly or after successively coarser text normalizations, keeping the earliest application in each duplicate group. This rule removes about 11 percent of records, leaving a universe of 7,064,256 patents. From the citation links we compute five-year forward citation counts, which are complete for application years through 2019 and set to missing thereafter.

The baseline sample keeps patents linked to a public firm, applied for in 1981–2018, with a valid non-negative IPO age and filed by firms with ≥ 50 such patents in the window. This leaves 1,788,076 patents from 1,286 firms. For analyses that require cluster assignments, we further restrict to patents with a non-noise cluster assignment (see Appendix A.2) and to firms with at least 50 such clustered patents. This leaves 876,332 patents from 895 firms. Note that to identify completed firm-cluster spells we exploit that we observe patenting activity outside the baseline window (1900–2024).

A.2 Embeddings and clustering methodology

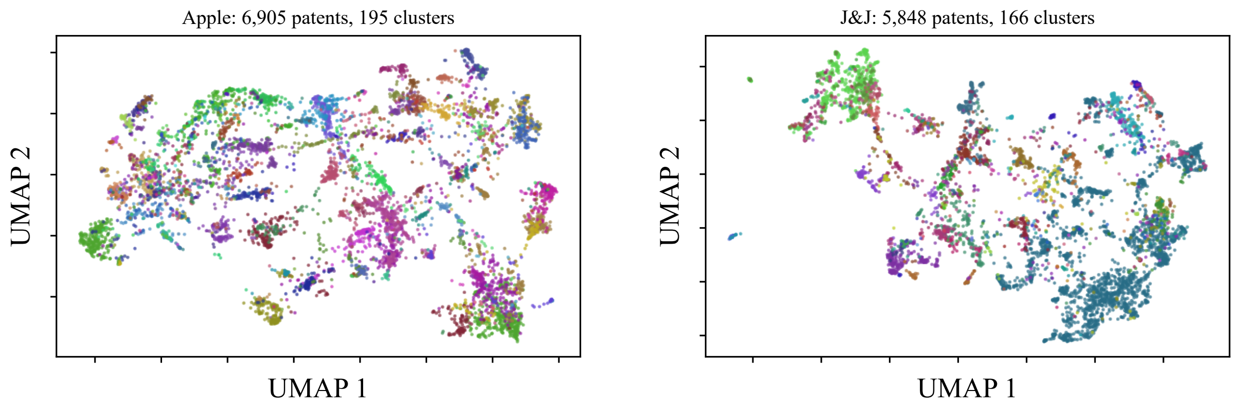
Embeddings. For each patent in the universe we concatenate the title and abstract and embed the text using OpenAI's `text-embedding-3-small` model at an output dimension of 256, yielding one vector per patent.

Density-based clustering within technology sections. We group patents by 1-digit CPC section, letting a patent whose codes span several sections enter each. Within each section we reduce the embeddings to 50 dimensions with UMAP and cluster the reduced points with HDBSCAN, a density-based algorithm that infers the number of clusters from the data and labels low-density points as noise. We fine-tune three hyperparameters—the UMAP neighborhood size (15–200), the HDBSCAN minimum cluster size (150–200), and its minimum sample count (0.7–1 times the minimum cluster size)—separately by section to maximize a standard density-based cluster validity index (DBCV), discarding configurations in which noise exceeds two-thirds of the section or a single cluster absorbs more than a quarter of it.

HDBSCAN initially labels roughly half of each section as noise. We reassign a noise patent to the cluster of its nearest assigned patent when its distance to that patent is below the cluster's median nearest-neighbor distance; more isolated patents remain noise. When a patent obtains candidate labels from several sections, we keep the label matching its first CPC code, then the one matching its modal CPC section; the residual case—fewer than 0.5 percent of patents—takes the candidate from the first section in a fixed ordering. This stage yields 3,441 fine clusters covering 3.7 million patents; 47 percent of patents remain noise.

Agglomeration to a common resolution. The fine clusters differ in granularity across sections. We use agglomerative clustering to anchor the granularity of our clusters in an external, widely used taxonomy. We target within each section the number of subclasses recorded in the patent classification system (CPC). We represent each fine cluster by the normalized mean embedding of its patents and, within each section, run average-linkage hierarchical clustering on the cosine distances between these centroids, stopping at as many clusters as the section has distinct CPC subclasses. The merge yields 672 clusters in the patent universe, of which 667 contain baseline-sample patents; these are the idea patches used throughout the paper. Noise patents are dropped only where an analysis requires a cluster assignment.

Visualization of embedding-based clusters. For illustration, Figure A.1 maps selected firms' patents into a two-dimensional embedding space and colors patents by cluster assignment.



(a) Apple

(b) Johnson & Johnson

Figure A.1: Embedding-based clusters for selected firms

Notes. Each point is a patent for Apple (panel (a)) or Johnson & Johnson (panel (b)) with a non-noise cluster assignment. UMAP is fit separately within each firm using patent embeddings; colors denote clusters.

A.3 Summary statistics

Table A.1 summarizes the patent-level innovation measures, constructed patent-distance variables, and inventor inputs used in the empirical analysis. Table A.2 reports firm-year moments for patenting, cluster entry, employment, sales, R&D, and R&D intensity in the public-firm panel. Table A.3 summarizes firm-cluster spells and global technology clusters, separately for the baseline and completed-spell samples. Figure A.2 shows pairwise correlations among the patent-level innovation measures.

A.4 Additional results

A.4.1 Robustness: diminishing returns

Figure A.3 repeats the relative-position slope estimates after varying the patch definition across firm × global embedding-cluster cells—our baseline—global embedding clusters, firm-×-CPC4 classes, and global CPC4 classes. The estimates remain negative across all six innovation outcomes and all four definitions, so the within-patch decline is not an artifact of the baseline embedding-cluster construction or of defining patches at the firm-specific level.

Figure A.4 estimates the same relative-position slopes separately within broad CPC technology sections; the decline appears across technology areas, with patent value again the weakest and noisiest outcome. Hence, the diminishing-returns pattern is not driven by one broad technology field or by the cross-sectional composition of clusters across fields.

Table A.1: Patent-level summary statistics

Variable	N	Mean	Std. dev.	p10	p50	p90
Panel A: patent-level innovation measures						
5-year citations	1,788,076	4.179	7.312	0.000	2.000	11.00
Patent value	1,788,076	11.74	21.64	0.055	4.361	29.63
Importance	1,446,529	0.256	0.209	0.006	0.236	0.523
Creativity	1,069,782	0.915	0.830	0.057	0.699	2.057
Panel B: constructed patent-level measures						
Minimum distance (5-year)	1,788,076	0.225	0.083	0.105	0.237	0.320
Novel combination	1,788,076	0.070	0.255	0.000	0.000	0.000
Panel C: inventor input						
Inventors on patent	1,787,949	2.743	1.746	1.000	2.000	5.000

Notes. The sample consists of public-firm patents with application years 1981–2018. Citations are five-year forward citations; patent value is in KPSS units; minimum distance is measured to the closest prior patent in the five-year backward window. Novel combination indicates at least one previously unobserved CPC-code pair, and inventor input counts inventors on the patent. Non-binary numeric variables are winsorized at the 1st and 99th percentiles; novel combination is not.

Table A.2: Firm-year summary statistics

Statistic	N	Mean	Std. dev.	p10	p50	p90
Patents per year	28,045	54.36	154.1	0.000	8.000	114.0
First-entry rate	17,753	0.270	0.410	0.000	0.123	0.800
Employees (thousands)	27,714	23.32	48.20	0.276	5.100	64.11
Sales	28,033	6,958	17,323	46.52	1,026	17,304
R&D expenditures	24,817	272.3	757.7	4.447	41.26	584.0
R&D intensity	24,722	0.259	0.902	0.008	0.058	0.314

Notes. The sample consists of CRSP-Compustat firm-years from 1981–2018 for firms in the public-firm patent baseline. Patent counts include zero-patent Compustat firm-years; missing R&D is not set to zero. The first-entry rate scales new cluster entries by lagged active clusters when the denominator is positive. Employees are in thousands; sales and R&D expenditures are in millions; R&D intensity is R&D/sales for positive sales. Non-binary numeric variables are winsorized at the 1st and 99th percentiles.

Table A.3: Cluster-level summary statistics

Measure	N	Mean	Std. dev.	p25	p50	p75	p90
Panel A: firm-cluster spells							
Baseline sample							
Patents	42,796	16.02	46.83	1.000	2.000	8.000	32.00
Patents, ≥ 2 patents	27,062	26.56	68.35	3.000	6.000	17.00	56.00
Active years	42,796	5.827	7.571	1.000	3.000	7.000	16.00
Active years, ≥ 2 patents	27,062	8.552	8.616	3.000	5.000	11.00	20.00
Span years	42,796	8.963	10.23	1.000	4.000	15.00	25.00
Span years, ≥ 2 patents	27,062	13.59	10.36	5.000	11.00	20.00	30.00
Patents per span year	42,796	1.367	1.926	0.632	1.000	1.000	2.197
Patents per span year, ≥ 2 patents	27,062	1.643	2.760	0.400	0.800	1.667	3.324
Mean 5-year citations	42,796	3.655	4.680	1.000	2.250	4.507	8.143
Mean 5-year citations, ≥ 2 patents	27,062	3.895	4.110	1.412	2.750	4.846	8.250
Completed spells							
Patents	28,434	5.178	11.22	1.000	2.000	4.000	11.00
Patents, ≥ 2 patents	14,834	9.638	18.05	2.000	4.000	8.000	21.00
Active years	28,434	2.852	3.224	1.000	1.000	3.000	7.000
Active years, ≥ 2 patents	14,834	4.562	3.850	2.000	3.000	6.000	10.00
Span years	28,434	5.626	7.091	1.000	1.000	8.000	17.00
Span years, ≥ 2 patents	14,834	9.897	7.748	4.000	8.000	14.00	21.00
Patents per span year	28,434	1.038	0.825	0.667	1.000	1.000	1.667
Patents per span year, ≥ 2 patents	14,834	1.120	1.391	0.375	0.667	1.333	2.000
Mean 5-year citations	28,434	3.796	5.257	1.000	2.000	4.800	9.000
Mean 5-year citations, ≥ 2 patents	14,834	4.193	4.704	1.286	2.750	5.300	9.500
Panel B: global technology clusters							
Baseline sample							
Patents	667	1,223	2,401	68.00	236.0	1,186	3,362
Firm-clusters	667	63.13	64.40	18.00	40.00	91.00	157.4
Completed spells							
Patents	667	257.4	456.2	32.00	88.00	259.5	660.8
Firm-clusters	667	41.96	39.85	13.00	28.00	60.00	101.2

Notes. The table summarizes public-firm patents assigned to non-noise clusters in application years 1981–2018. Completed spells require observed entry and last patent year within 1981–2018 and exclude right-censored firm-cluster cells in the extended patent history. Active years count years with at least one patent; span years count first-to-last baseline patent years. Outcomes are winsorized at the 1st and 99th percentiles.

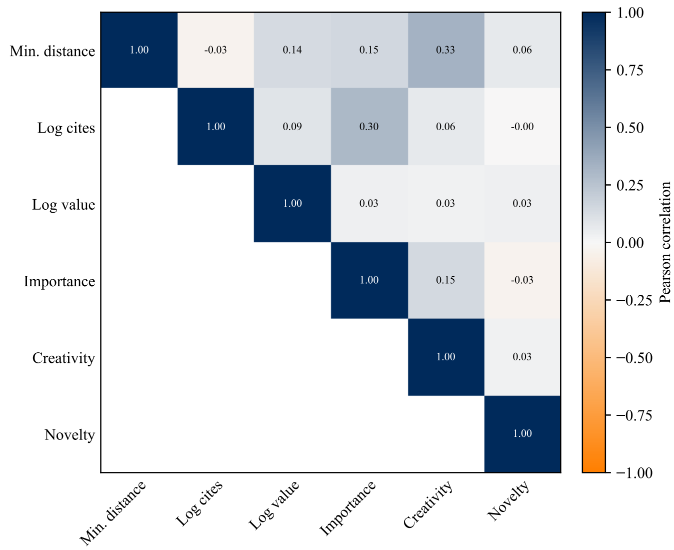


Figure A.2: Patent-measure correlations

Notes. The figure reports pairwise Pearson correlations for patent-level measures in the public-firm baseline sample. Citations and patent value are log transformed; continuous measures are winsorized at the 1st and 99th percentiles, while novel combination remains binary. Correlations use pairwise nonmissing observations.

A.4.2 R&D input proxies

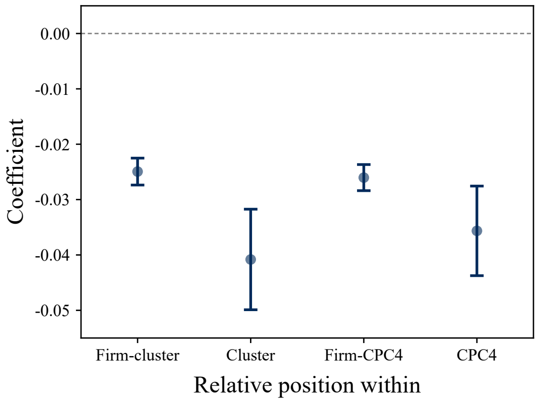
A natural concern is that the within-patch decline in patent quality could reflect declining research inputs rather than diminishing returns to the patch itself. Since the analysis is at the patent level, this concern requires that input *per patent* falls as the firm advances through the patch. Figure A.5 shows that it does not: the number of inventors per patent, a simple proxy for research input, is roughly flat over the firm-cluster sequence. Successive patents represent comparable input for diminishing output.

A.4.3 Post-entry innovation ratios

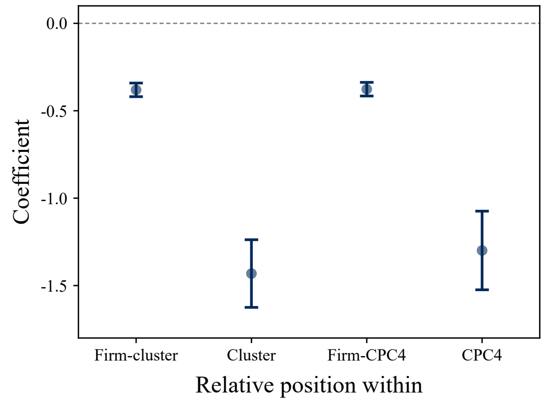
Table A.4 compares early patents after first entry into a firm-cluster with the firm's own prior-year innovations outside that cluster. The averages are often above one for creativity, citations, and especially patent value. At the same time, medians are mostly below one, suggesting that the post-entry boost is concentrated in a right tail of valuable clusters rather than a uniform shift across entries.

A.4.4 Inter-cluster waiting times

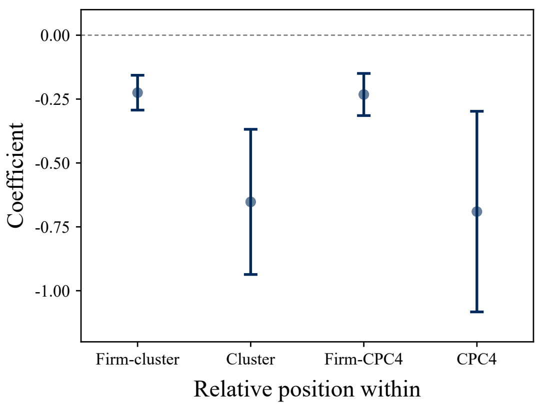
We use a variety of filters to isolate episodes when a firm most likely acts as one sequential line. They trade off how closely they align with the model's notion of a sequential search versus how representative the underlying sample is. There are two groups of filters. *Quiet-period* filters



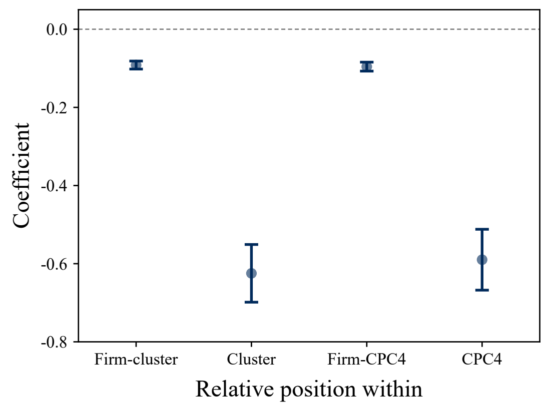
(a) Minimum distance



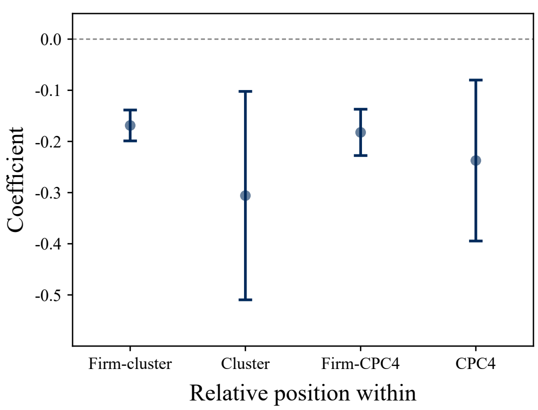
(b) 5-year citations



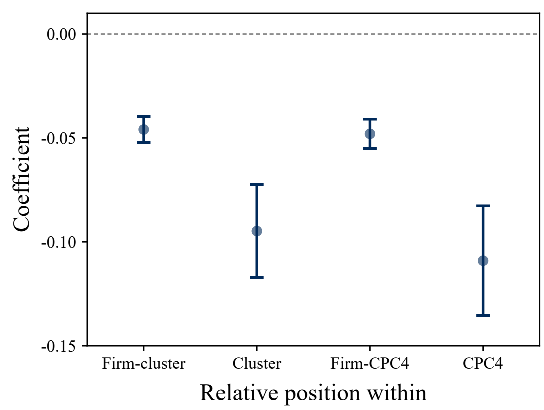
(c) Patent value



(d) Breakthrough importance



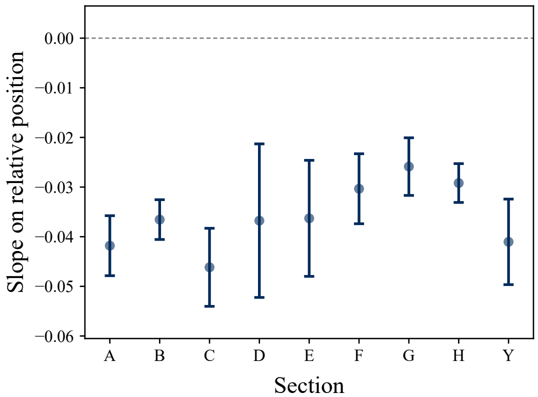
(e) Creativity



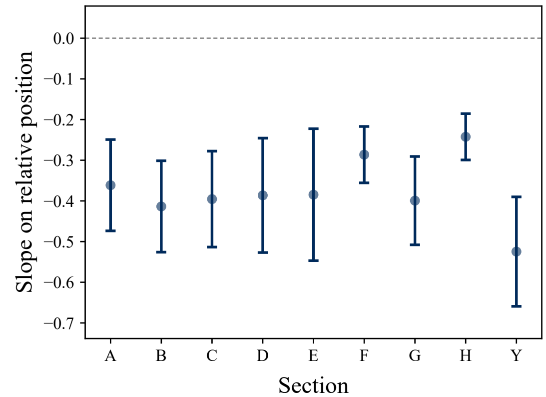
(f) Novel combinations

Figure A.3: Relative-position robustness across patch definitions

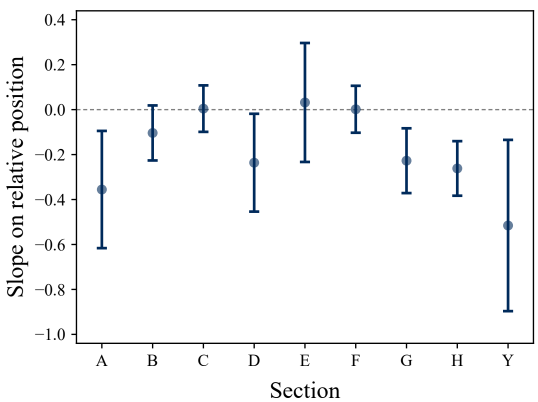
Notes. Points report relative-position slopes with 95% confidence intervals. The sample is the public-firm baseline restricted to non-singleton firm-patch cells. Relative position is recomputed for each patch definition. The embedding-cluster specifications absorb application-year, firm, and embedding-cluster fixed effects, with standard errors clustered by embedding cluster. The CPC4 specifications absorb application-year, firm, and CPC4 fixed effects, with standard errors clustered by CPC4. Continuous outcomes are winsorized at the 0.5th and 99.5th percentiles; novel combination is binary and is not winsorized. Importance and creativity use 1981–2016; the other panels use 1981–2018.



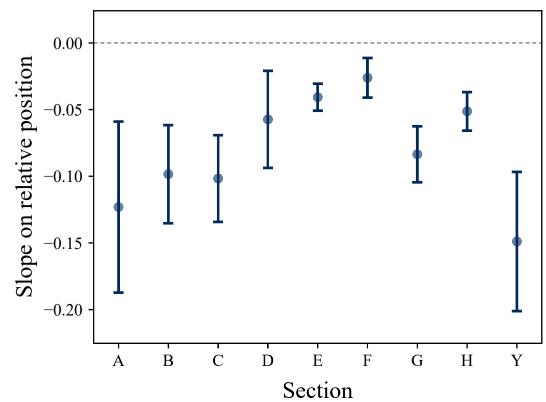
(a) Minimum distance



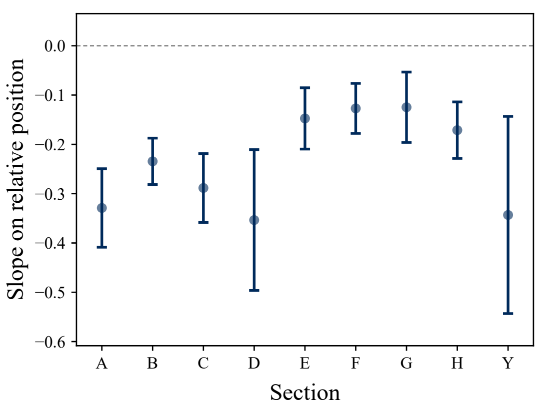
(b) 5-year citations



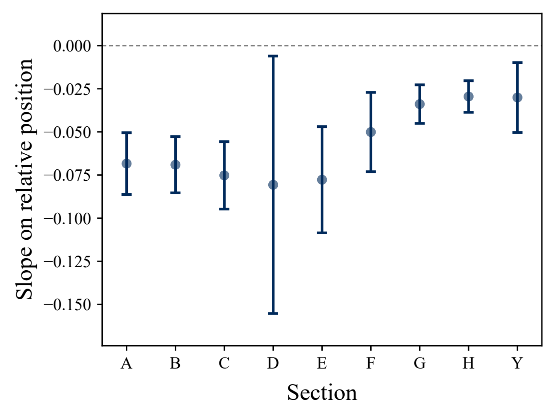
(c) Patent value



(d) Breakthrough importance



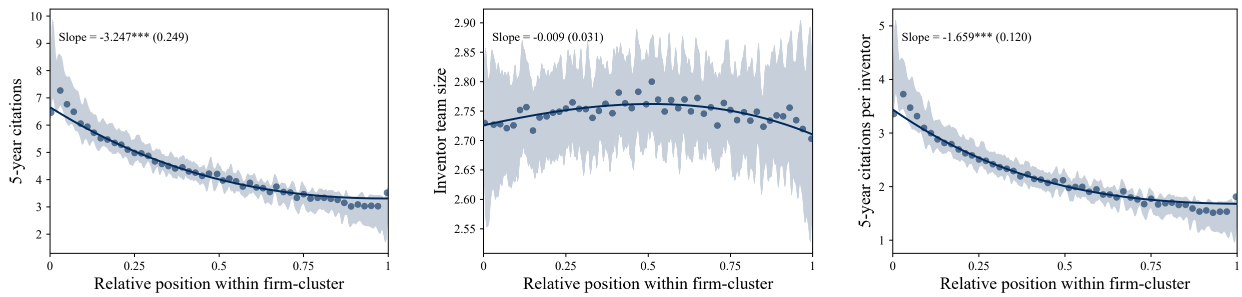
(e) Creativity



(f) Novel combinations

Figure A.4: By-CPC-section relative-position slopes

Notes. Points report relative-position slopes with 95% confidence intervals. The sample is public-firm baseline patents restricted to non-singleton firm-clusters. Relative position scales each firm-cluster sequence from 0 to 1. Specifications absorb application-year and firm-cluster fixed effects. Standard errors are clustered by agglomerative technology cluster. Continuous outcomes are winsorized at the 0.5th and 99.5th percentiles. Importance and creativity use 1981–2016; the other panels use 1981–2018.



(a) Citation output **(b)** Inventor team size **(c)** Citations per inventor

Figure A.5: R&D inputs and efficiency over relative position

Notes. The sample is public-firm baseline patents assigned to non-singleton firm-clusters. Relative position scales each firm-cluster sequence from 0 to 1. The specification absorbs application-year and firm-cluster fixed effects. Citation output and citations per inventor use raw 5-year forward citations.

Table A.4: Post-entry innovation ratios

	Creativity		Citations		Importance		Patent value	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Panel A: all complete-lifecycle firm-clusters								
Ratio (1st 10 patents)	1.12	0.86	1.02	0.54	1.00	0.98	1.22	0.93
Ratio (1st 10% patents)	1.22	0.90	1.10	0.50	1.01	0.99	1.07	0.88
Panel B: complete-lifecycle firm-clusters with calendar span ≥ 10 years								
Ratio (1st 10 patents)	0.98	0.83	1.08	0.72	0.98	0.96	1.71	1.08
Ratio (1st 10% patents)	1.15	0.92	1.25	0.73	1.03	1.00	1.18	0.91

Notes. The sample is the public-firm baseline restricted to firm-cluster patents whose first observed entry year and last observed patent year both fall within 1981–2018. For each firm-cluster entry, the numerator is the mean innovation measure among the first 10 post-entry patents, or among the first 10% of post-entry patents. The denominator is the firm's mean innovation measure in the year before entry, excluding patents in the focal cluster. The bottom panel restricts the sample to an entry-to-exit calendar span of at least 10 years. Innovation measures are winsorized at the 0.5th and 99.5th percentiles before ratios are formed.

count the wait only when the firm goes fully dark, i.e., it has no patent in any cluster, between leaving one cluster and entering the next. This is the closest analogue to the model's notion of waiting during search. *Per-spell* filters count the wait from the end of one cluster activity to the next entry into a different cluster, without requiring full idleness in between. This approach maps less directly to the model but expands the sample.

- q-single (quiet): the firm holds one cluster, all prior engagements are finished, it goes fully dark, then enters one new cluster. This is the strictest single-line analogue.
- prev-single (per-spell): the prior engagement had at most one cluster active during its span; the firm then enters a new cluster, not necessarily after going fully dark.
- q-completed (quiet): like q-single, but the firm may have held several clusters before going dark, as long as all are finished. Relaxes "one prior cluster" to "all prior clusters finished."
- q-all (quiet): any fully-dark window, regardless of how many clusters preceded it or whether they finished. This is the loosest quiet filter.
- low-parallel (per-spell): allows up to two clusters active during the prior engagement and at most one during the gap. The most permissive filter we report.

We construct the wait time on two samples. For other moments, we use the sample of public firms with 50+ patents. But since this gives relatively thin counts, we also look at the sample of *all* corporate assignees with 50+ patents. Indeed, we treat the latter as the baseline, as clean single-line transitions are common enough there to provide adequate precision. Reassuringly, the two samples yield similar statistics.

Table A.5 summarizes the measured wait times along with the implied value for λ given the calibrated value of $\delta_F = 0.0548$. Across all approaches, the implied value of λ falls into a relatively narrow band of approximately 0.15–0.2.

A.4.5 Time-series evidence on cluster duration and entry

This section provides methodological details about and robustness checks on the time-series evidence presented in the growth-slowdown analysis of Section 4.3. Both statistics described there—how long a firm remains active in a cluster once it has entered, and how often firms enter clusters new to them—are vulnerable to truncation effects that could masquerade as trends. Throughout, the unit is a firm-cluster pair in the public 50+ baseline. We track firm-cluster activity outside the 1981–2018 baseline window as well—in practice chiefly the 1970s and 2019–2024—so that entry dates are not left-censored at 1981 and ten-year follow-up is complete for entry cohorts through 2015. Cluster spells require at least two patents over their observed history.

Table A.5: Inter-cluster waiting times

		All corporate			Public 50+		
Approach	Restriction	N	Wait	λ	N	Wait	λ
q_single *(baseline)*	fully dark; prior = 1 cluster	637	4.90	0.149	88	4.19	0.184
prev_single	prior engagement ≤1 cluster active	597	4.89	0.150	83	4.13	0.187
q_completed	fully dark; prior all finished (≥1 cluster)	955	5.13	0.140	132	4.63	0.161
q_all	any quiet window	966	5.10	0.141	135	4.57	0.164
low_parallel	≤2 clusters active (comparison)	1,669	4.68	0.159	223	4.21	0.183

Notes. The all-corporate sample contains corporate assignees with at least 50 patents; the public 50+ sample matches the public-firm baseline sample. All rows require at least two patents per firm-cluster, impose a one-year minimum search gap, and use one spell per firm-cluster. Wait is in years. The reported λ maps each wait into $\lambda = 1/\text{wait} - \delta_F$, where $\delta_F = 0.0548$ is the same firm exit rate used in the calibration.

Duration. A raw first-to-last measure of patent duration can mechanically shorten for recent cohorts. We therefore measure duration on a window of fixed length. We consider a firm-cluster entered in year t_e—dated by the firm's first patent there or, to guard against stray single-patent assignments, by the peak-flow rule of footnote 12—and count the years with at least one patent among the H years from t_e, and average across episodes by entry cohort (the columns of Panel A). The baseline is $H = 10$, with the last cohort ending in 2015 so that every episode has a complete window. Table A.6 varies the horizon, with $H = 5$ extending coverage to cohorts through 2018, and repeats the exercise on the all-corporate 50+ sample; the pattern is unchanged in both cases, showing it is neither horizon-specific nor specific to public firms.

Entry. The second observable is the rate at which firms enter clusters new to them, per active firm-year. The first-ever entry rate is subject to a somewhat mechanical drift: a firm can enter a cluster for the first time only once, so the rate falls as firms age. This matters because the average patenting age of active sample firms roughly doubles over the sample period. We therefore also consider two alternative measures. The entry/re-entry rate also counts returns after more than five years without a patent in the cluster. In Table 7, Panel C compares firms at the same "age," where age here is based on dating firm birth by the first clustered patent in our baseline sample, and computes rates within the first five years of the firm's patenting life. At fixed firm age the decline largely disappears, which is why the main text treats the entry evidence as suggestive. Table A.6 adds rates without the two-patent filter and an alternative normalization by active firm-cluster-years.

Table A.6: Robustness for firm-cluster persistence and entry over time

Panel A. Duration robustness

Definition	Sample	1981–1989	1990–1999	2000–2009	2010–2015	2010–2018
First-patent date; H=10	Public 50+	3.16	3.97	3.68	3.56	
First-patent date; H=5	Public 50+	2.01	2.47	2.33		2.42
First-patent date; H=10	All corporate 50+	3.32	3.68	3.69	3.52	

Panel B. Entry-rate robustness

Event	Definition	Denom.	Sample	1981–1989	1990–1999	2000–2009	2010–2018
First-ever entry	$\geq$ 2 patents	Firm-year	Public 50+	1.43	1.32	1.10	0.75
First-ever entry	$\geq$ 2 patents	Cluster-year	Public 50+	0.13	0.13	0.10	0.06
First-ever entry	All entries	Firm-year	Public 50+	2.21	1.97	1.81	1.53
First-ever entry	$\geq$ 2 patents	Firm-year	All corporate 50+	0.96	0.89	0.84	0.66
First entry or re-entry	>5-year return; $\geq$ 2 patents	Firm-year	Public 50+	2.13	2.03	1.81	1.73
First entry or re-entry	>5-year return; $\geq$ 2 patents	Cluster-year	Public 50+	0.19	0.20	0.16	0.14
First entry or re-entry	>5-year return; $\geq$ 2 patents	Firm-year	All corporate 50+	1.25	1.25	1.17	1.16

Panel C. Fixed-age entry robustness

Event	Definition	Denom.	Sample	1982–1989	1990–1999	2000–2009
First-ever entry	First 5 patenting years; $\geq$ 2 patents	Firm-year	Public 50+	1.44	1.79	1.47
First-ever entry	First 5 patenting years; $\geq$ 2 patents	Cluster-year	Public 50+	0.46	0.46	0.47
First-ever entry	First 10 patenting years; $\geq$ 2 patents	Firm-year	Public 50+	1.39	1.59	1.42
First-ever entry	First 10 patenting years; $\geq$ 2 patents	Cluster-year	Public 50+	0.31	0.28	0.28

Notes. Definitions follow Table 7. H is the duration follow-up window, equal to 10 in the main text; columns are entry cohorts in Panel A, calendar years in Panel B, and firm birth cohorts in Panel C. Panel A cohorts end so that every episode has a complete window ($H = 5$ through 2018, $H = 10$ through 2015). In Panel B the denominator is active firm-years or active firm-cluster-years (clusters with at least one patent by the firm that year); "all entries" drops the two-patent requirement. The all-corporate 50+ sample comprises corporate assignees with at least 50 clustered patents and at least two clusters. Panel C fixes firm age by birth cohort; 2010s cohorts are too thin and are omitted.

Appendix B Appendix: Theory

B.1 Quality-weight accounting on the BGP

This appendix derives the quality-weight objects used in Section 3.2.3. It also explains why the cross-sectional distribution can be summarized by two backward equations and one linear recursion.

Within-spell quality weights. Let $z = (q/\bar{Q})^{\sigma-1}$ denote the average-quality weight. During an exploitation spell, let

$$z_t = z_{\text{entry}} M_t,$$

where z_{entry} is the quality weight at the start of the spell and

$$M_t = \exp\{-(\sigma-1)gt\} \prod_{i:T_i \le t} \theta(X_{T_i})^{\sigma-1}$$

is the within-spell multiplier from patents net of aggregate dilution. The process M_t depends on the current patch path and patent arrivals. It does not depend on the firm's inherited z_{entry}, because each exploitation spell starts from a fresh draw of X_0. Hence within-spell dynamics and inherited quality weight factor as

$$E[z_{\text{entry}} M_t] = E[z_{\text{entry}}]E[M_t]. \tag{B.1}$$

Two backward equations. Define

$$\Psi^{\text{surv}}(X) \equiv E[M_{\tau^I} \mathbf{1}\{\text{non-death exit}\} \mid X_0 = X], \qquad \Xi(X) \equiv E\left[\int_0^{\tau^I} M_t \, dt \mid X_0 = X\right].$$

The first object is the quality-weight multiplier carried from exploitation into exploration, with firm death counted as zero. The second is the cumulative quality weight generated while exploiting a patch. By the Feynman–Kac formula,

$$\frac{\sigma_x^2}{2}(\Psi^{\text{surv}})'' + \mu(\Psi^{\text{surv}})' + \left[\gamma(\theta(X)^{\sigma-1} - 1) - (\sigma-1)g - (\delta + \delta_F)\right]\Psi^{\text{surv}} + \delta = 0, \tag{B.2}$$

$$\frac{\sigma_x^2}{2}\Xi'' + \mu\Xi' + \left[\gamma(\theta(X)^{\sigma-1} - 1) - (\sigma-1)g - (\delta + \delta_F)\right]\Xi + 1 = 0. \tag{B.3}$$

The boundary conditions are

$$\Psi^{\text{surv}}(x^*) = 1, \qquad \Xi(x^*) = 0,$$

together with boundedness on the upper tail. The source term in (B.2) is δ, not $\delta + \delta_F$, because creative destruction ends the active spell but the firm survives; firm death contributes zero to the non-death-exit indicator. Averaging across initial patch types gives

$$\bar{\Psi}^{\text{surv}} = \sum_{k=1}^{K} p_k \Psi^{\text{surv}}(x_{0,k}), \qquad \bar{\Xi} = \sum_{k=1}^{K} p_k \Xi(x_{0,k}).$$

Exploitation-entry weights. The flow into exploitation is $\zeta = F/\bar{T}$. Of this flow, $\delta_F F$ firms are new entrants. Hence a fraction $\delta_F \bar{T}$ of exploitation entries are outside entrants with quality weight $\bar{\Lambda}_E^{\sigma-1}$. The remaining entries are surviving incumbents.

A surviving incumbent's expected quality-weight multiplier between two exploitation entries is

$$\bar{\Psi}^{\text{surv}} \cdot \Lambda^{\sigma-1} \cdot \frac{\lambda}{\lambda + \delta_F + (\sigma - 1)g}.$$

The first term is accumulated during exploitation. The second is the quality scaling on the next patch. The third combines the arrival rate of discovery during exploration with dilution at rate $(\sigma - 1)g$ and death at rate δ_F.

Therefore the average quality weight at exploitation entry satisfies

$$\bar{z}_{\text{entry}} = \delta_F \bar{T}\, \bar{\Lambda}_E^{\sigma-1} + \bar{z}_{\text{entry}} \frac{\lambda \Lambda^{\sigma-1} \bar{\Psi}^{\text{surv}}}{\lambda + \delta_F + (\sigma - 1)g}.$$

Solving yields (24). The denominator is positive exactly when the contractivity condition (27) holds.

Exploiting and exploring firms. For exploiting firms, factorization gives total quality-weight time in exploitation equal to $\bar{z}_{\text{entry}} \bar{\Xi}$ per spell. Dividing by the unweighted spell duration gives

$$\bar{z}_I = \bar{z}_{\text{entry}} \frac{\bar{\Xi}}{\bar{\tau}^I},$$

which is (25).

For exploring firms, condition first on reaching exploration. The expected quality weight at the start of exploration is

$$\bar{z}_{\text{entry}} \frac{\bar{\Psi}^{\text{surv}}}{1 - \delta_F \bar{\tau}^I}.$$

Exploration duration is exponential with rate $\lambda + \delta_F$, while quality weight is diluted at rate

$(\sigma - 1)g$. The cross-sectional average dilution factor is therefore

$$\frac{\lambda + \delta_F}{\lambda + \delta_F + (\sigma - 1)g}.$$

Multiplying these terms gives (26).

B.2 Derivation of the average-quality CES closure

By definition of average product quality,

$$\frac{1}{\mathcal{N}} \int_0^{\mathcal{N}} \left(\frac{q_j}{\bar{Q}} \right)^{\sigma-1} dj = 1.$$

Let Z_A denote the total z-weight of active products and Z_L the total z-weight of legacy products. The BGP normalization is

$$Z_A + Z_L = \mathcal{N}.$$

Active products. The flow into exploitation is $F/\bar{T}$. Each entering spell has expected entry weight $\bar{z}_{\text{entry}}$ and cumulative within-spell weight $\bar{\Xi}$. Hence

$$Z_A = \frac{F}{\bar{T}} \bar{z}_{\text{entry}} \bar{\Xi} = \frac{\mathcal{N}}{\chi_{\mathcal{N}} \bar{T}} \bar{z}_{\text{entry}} \bar{\Xi}. \tag{B.4}$$

Legacy products. Legacy products are created when firms voluntarily leave an active patch. The quality weight flowing into legacy is the quality weight at non-death exit net of spells that end by creative destruction. Since creative destruction arrives during exploitation at rate δ, this flow is

$$\frac{F}{\bar{T}} \bar{z}_{\text{entry}} \big(\bar{\Psi}^{\text{surv}} - \delta \bar{\Xi} \big).$$

Legacy weights then decay through aggregate dilution, creative destruction, and firm death, at total rate $(\sigma - 1)g + \delta + \delta_F$. Therefore

$$Z_L = \frac{\mathcal{N}}{\chi_{\mathcal{N}} \bar{T}} \bar{z}_{\text{entry}} \frac{\bar{\Psi}^{\text{surv}} - \delta \bar{\Xi}}{(\sigma - 1)g + \delta + \delta_F}. \tag{B.5}$$

Normalization. Substituting (B.4) and (B.5) into $Z_A + Z_L = \mathcal{N}$ and dividing by $\mathcal{N}$ gives

$$\frac{\bar{z}_{\text{entry}}}{\chi_{\mathcal{N}} \bar{T}} \left[\bar{\Xi} + \frac{\bar{\Psi}^{\text{surv}} - \delta \bar{\Xi}}{(\sigma - 1)g + \delta + \delta_F} \right] = 1.$$

The bracket equals

$$\frac{[(\sigma-1)g+\delta_F]\bar{\Xi}+\bar{\Psi}^{\mathrm{surv}}}{(\sigma-1)g+\delta+\delta_F},$$

which proves (28).

B.3 Flow representation of average-quality growth

The equilibrium closure condition also has a flow interpretation in average-quality weights. Let

$$\Phi_z(x)\,dx \equiv \bar{Q}^{-(\sigma-1)} \sum_{\{j\in A:\, X_j\in[x,x+dx]\}} q_j^{\sigma-1}$$

denote the stationary density of active-product z-weight across patch densities. In steady state, Φ_z solves the forward equation

$$\begin{aligned} 0 = -\mu\,\Phi_z'(x) + \frac{\sigma_x^2}{2}\Phi_z''(x) + \Big[\gamma\big(\theta(x)^{\sigma-1}-1\big) - \delta - \delta_F - (\sigma-1)g\Big]\Phi_z(x) \\ + \sum_k Z_k^{\Phi}\,\delta_{\mathrm{Dirac}}(x-x_{0,k}), \end{aligned} \tag{B.6}$$

with absorbing boundary $\Phi_z(x^*)=0$ and the natural upper-tail condition $\Phi_z(x)\to 0$ as $x\to\infty$.[B.1] The source term at the initial patch density is

$$Z_k^{\Phi} = p_k\left[\lambda\Lambda^{\sigma-1}Z_E + \delta_F F\bar{\Lambda}_E^{\sigma-1}\right], \qquad Z_E = F^E\bar{z}_E. \tag{B.7}$$

The first term is the z-weight of discoveries by exploring incumbents, and the second is the z-weight of replacement entrants. The absorbing boundary moves active products into the legacy stock; it is not an additional source of aggregate quality creation.

Integrating (B.6) together with the stationary legacy-stock equation gives the flow representation of the CES normalization. Normalizing by the actual variety mass $\mathcal{N}$, this representation is

$$(\sigma-1)g = \frac{1}{\mathcal{N}}\int \gamma\big[\theta(x)^{\sigma-1}-1\big]\Phi_z(x)\,dx + \lambda\Lambda^{\sigma-1}\frac{f_E}{\chi\mathcal{N}}\bar{z}_E + \frac{\delta_F}{\chi\mathcal{N}}\bar{\Lambda}_E^{\sigma-1} - (\delta+\delta_F). \tag{B.8}$$

Equation (B.8) is the forward-KFE representation of the accounting behind (28). While useful for interpretation, we use the stock condition (28) in the definition and numerical computation of equilibrium, because it can be computed from the backward BVP objects without solving for

[B.1] With deterministic depletion ($\sigma_x=0$), the equation is first order and the absorbing condition is replaced by the upwind boundary flux $|\mu|\,\Phi_z(x^*)$, the flow of quality weight into the legacy stock.

Φ_z.

B.4 Sufficient conditions for g to collapse to g_q

This section provides a proof of Proposition 1.

Extended model. To make the "no legacy products" requirement a clean condition on primitives, we introduce a legacy-specific obsolescence rate $\delta_L \geq 0$ on top of creative destruction δ and firm death δ_F. Hence, a legacy line (i.e., a product a firm has voluntarily abandoned but still sells) faces total hazard $\delta + \delta_F + \delta_L$; an active line (under exploitation) still faces only $\delta + \delta_F$. The baseline model sets $\delta_L = 0$; the no-legacy benchmark is $\delta_L \to \infty$.

Parameter δ_L enters the equilibrium system through four objects. In the legacy value (9), the value-matching condition (12), and the CES growth closure (28), the only change is the substitution of the legacy hazard: $\hat{W}_0 = 1/(\hat{\rho} + \delta + \delta_F + \delta_L)$, and the legacy-decay rate in the closure becomes $(\sigma - 1)g + \delta + \delta_F + \delta_L$, so that

$$\frac{\bar{z}_{\text{entry}}}{\chi_{\mathcal{N}} \bar{T}} \left[\bar{\Xi} + \frac{\bar{\Psi}^{\text{surv}} - \delta \bar{\Xi}}{(\sigma - 1)g + \delta + \delta_F + \delta_L} \right] = 1.$$

The variety ratio requires more care, because active and legacy lines now die at different rates and must be tracked separately. There is one active line per exploiting firm, $\mathcal{N}_A = F^I$. Legacy lines are created only by voluntary exit, which ends a spell with probability $p_{\text{vol}} = 1 - (\delta + \delta_F)\bar{\tau}^I$ (by the same occupation-time argument as $p_{\text{death},I} = \delta_F \bar{\tau}^I$), so their inflow is $(F/\bar{T})\, p_{\text{vol}}$ and their stationary stock is $\mathcal{N}_L = F\, p_{\text{vol}}/[\bar{T}(\delta + \delta_F + \delta_L)]$. Hence

$$\chi_{\mathcal{N}}(\delta_L) = f_I + \frac{1 - (\delta + \delta_F)\bar{\tau}^I}{\bar{T}\,(\delta + \delta_F + \delta_L)}.$$

At $\delta_L = 0$ this reduces to (23), using the stationarity identity $\lambda f_E + \delta_F = 1/\bar{T}$; as $\delta_L \to \infty$, $\chi_{\mathcal{N}} \to f_I$, i.e., $\mathcal{N} = F^I$, which is the form used in the proof below. All other equilibrium equations are unchanged in form; δ_L affects them only indirectly through x^* (and through the extended $\chi_{\mathcal{N}}$ where it appears).

Proof of Proposition 1

Proof. With $\delta_F = 0$, no firm dies and no entry occurs, so the economy is a fixed mass F of firms cycling independently through exploitation and exploration. As log-quality gains are additive and each cycle starts from a fresh draw of initial patch density, the per-cycle gains and durations (G_n, T_n) are i.i.d. across cycles and firms, with finite means $E[G] = \bar{h} + \log \Lambda$ and

$E[T] = \bar{\tau}^I + 1/\lambda = \bar{T}$. By the renewal–reward theorem, $\ln q_j(t)/t \to E[G]/E[T] = g_q$ almost surely for each firm j; it remains to show that the level of the index grows at exactly this rate on the balanced growth path.

With $\delta_L \to \infty$, legacy products vanish instantly, so the varieties in the index are the active lines, one per exploiting firm: $\mathcal{N} = F^I = f_I F$ with $f_I = \bar{\tau}^I/\bar{T}$ by Little's law. With $\sigma \to 1$, the CES average-quality index reduces to the geometric mean,

$$\ln \bar{Q}(t) = \frac{1}{F^I} \int_{j \in \mathcal{A}(t)} \ln q_j(t)\, dj,$$

where $\mathcal{A}(t)$ is the set of exploiting firms. By the exact law of large numbers, given a continuum of independent firms, the flows and cross-sectional averages below are deterministic and equal their population expectations.

Let $S(t) \equiv \int_{j \in \mathcal{A}(t)} \ln q_j(t)\, dj = F^I \ln \bar{Q}(t)$. On a BGP, cross-sectional conditional means of $\ln q$ grow linearly at a common rate a. The stock S changes for three reasons. First, exploiting firms gain log quality through patents: by the same renewal-reward accounting used for (19) and (25), the aggregate flow of these gains equals the flow of spell starts, $\zeta = F/\bar{T}$, times the expected gain per spell, $\bar{h}$. Second, spells end (by hitting the threshold or creative destruction) at flow ζ, removing firms whose mean log quality at exit is $e(t)$. Third, searchers re-enter at flow $\lambda F^E = \zeta$; a firm entering at t exited at $t - s$ with $s \sim \mathrm{Exp}(\lambda)$ and carries its exit log quality plus $\log \Lambda$, so the mean log quality at entry is $E_s[e(t-s)] + \log \Lambda = e(t) - a/\lambda + \log \Lambda$, using the linearity of $e(\cdot)$. The exit and entry means therefore cancel up to the search delay:

$$\dot{S}(t) = \zeta\, \bar{h} + \zeta \Big(\log \Lambda - \frac{a}{\lambda} \Big).$$

Since $S(t) = F^I \ln \bar{Q}(t)$ grows at rate $F^I a$ and $F^I = \zeta\, \bar{\tau}^I$,

$$a\, \bar{\tau}^I = \bar{h} + \log \Lambda - \frac{a}{\lambda} \qquad \Longleftrightarrow \qquad a = \frac{\bar{h} + \log \Lambda}{\bar{\tau}^I + 1/\lambda} = \frac{E[G]}{E[T]} = g_q.$$

Hence $\ln \bar{Q}(t) = g_q t + c$ on the balanced growth path, giving $g = g_q$. □

B.5 The vintage composition of patenting

Denote the exploitation duration for type k: $T_k = (x_{0,k} - x^*)/|\mu|$ and the survival probability $P_k = e^{-\delta_{\text{exit}} T_k}$, with $\delta_{\text{exit}} = \delta + \delta_F$.

Proposition B.1. *On the BGP with $\sigma_x = 0$, the share of patents from new clusters over horizon h is*

$$S_{count}(h) = 1 - \frac{1}{h} \cdot \frac{\sum_k p_k I_k(h)}{\sum_k p_k(1-P_k)}$$

where

$$I_k(h) = \int_0^{\min(h,T_k)} \left(e^{-\delta_{exit}t} - P_k\right) dt = \begin{cases} \dfrac{1-e^{-\delta_{exit}h}}{\delta_{exit}} - P_k h & h \le T_k \\ \dfrac{1-P_k}{\delta_{exit}} - P_k T_k & h > T_k. \end{cases}$$

$S_{count}(0) = 0$, $S_{count}(\infty) = 1$, S_{count} *is increasing and concave.*

Proof. On the BGP, total patent flow is constant (γF^I). The pre-existing contribution at time t is $\gamma F^I \cdot P_{\text{surv}}(t)$, where $P_{\text{surv}}(t)$ is the fraction of exploiting firms still on their reference-date cluster. Since γF^I cancels in the ratio, everything reduces to P_{surv}. The cross-sectional age density of type-k exploiting firms is the truncated exponential $f_{s|k}(s) = \delta_{\text{exit}} e^{-\delta_{\text{exit}} s}/(1-P_k)$ for $s \in [0, T_k]$. A firm at age s survives on its cluster to time t iff $s + t < T_k$ and no exit shock hits in $[0, t]$. The type-k cross-sectional mass is proportional to $p_k(1-P_k)/\delta_{\text{exit}}$. Integrating, the $(1-P_k)$ factors cancel:

$$P_{\text{surv}}(t) = \frac{\sum_k p_k(e^{-\delta_{\text{exit}}t} - P_k)^+}{\sum_k p_k(1-P_k)}$$

Then $S_{\text{count}}(h) = 1 - \frac{1}{h}\int_0^h P_{\text{surv}}(t)\, dt$ yields the result. □

This formula depends on parameters only through the *endogenous* cluster durations $\{T_k\}$, via x^*—which we can approximately characterize in closed form—and the exit rate δ_{exit}.

With a single patch type and $\delta_{\text{exit}} = 0$, the formula collapses to:

$$S_{\text{count}}(h) = \begin{cases} h/(2T) & h \le T \\ 1 - T/(2h) & h > T \end{cases}$$

where $T = (x_0 - x^*)/|\mu|$. So the half-life—the horizon at which half of patents come from new clusters—is $h_{1/2} = T$, i.e., exactly the cluster exploitation duration.

The comparative statics of S_{count} (in partial equilibrium, i.e., holding (g, L^P) fixed) follow from $\partial S_{\text{count}}/\partial T < 0$—shorter clusters raise the new-cluster share—and the analytical derivatives for x^*. Since $T_k = (x_{0,k} - x^*)/|\mu|$: $\frac{\partial T_k}{\partial \vartheta} = -\frac{1}{|\mu|}\frac{\partial x^*}{\partial \vartheta}$ for any parameter ϑ that enters only through x^* (i.e., all parameters except $|\mu|$ and x_0).

- Higher $|\mu|$ *raises* S_{count} directly, as cluster duration is shorter.
- Higher λ raises x^* ($\partial x^*/\partial \lambda > 0$), as the outside option is more valuable, which *raises* S_{count}.

- A higher γ or a higher ϕ (indexing $\theta(X)$) *lowers* S_{count}. The intuition is that patents are more frequent or more valuable, so the firm tolerates more depletion and stays longer.