A Model of Knowledge Creation and Diffusion*

Sichuang Xu† Shengxing Zhang‡

March 2026

Abstract

We develop a dynamic general equilibrium model of knowledge creation and diffusion across heterogeneous firms. Firms invest in R&D and in firm-specific inputs (data and software) that raise the productivity of idea creation but reduce the transferability of new ideas. An extreme-value structure delivers a tractable law of motion for the technological frontier and implies that a fall in the cost of firm-specific inputs can increase innovative effort yet slow aggregate growth by weakening diffusion. Using U.S. patents matched to Compustat, we show that software accumulation by leading firms predicts slower diffusion: a 1% increase in leaders' software stock reduces upward citations from follower firms by 0.08%, while changes in leaders' R&D have positive effect. Calibrated to these moments, the model reproduces the post-2000 decline in vertical knowledge diffusion and quantifies how falling software costs can lower long-run growth.

Keywords: Knowledge diffusion; secrecy; intangible capital; software; innovation; growth
JEL codes: O31; O33; O34; D83

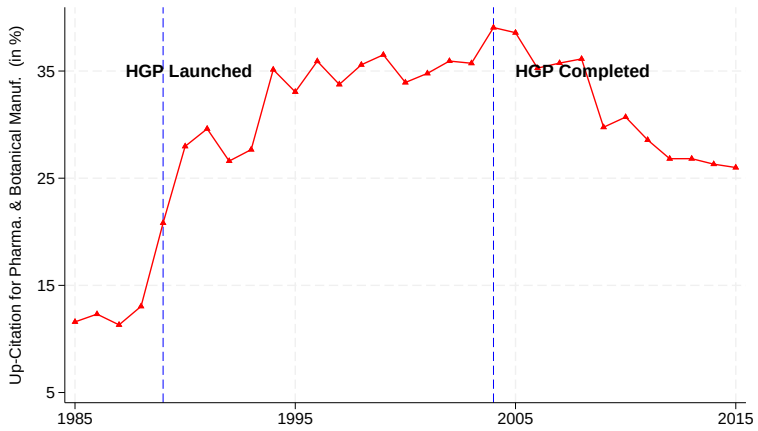
*We are grateful to participants at the EFG session of the 2023 Barcelona Summer Forum and the first in-person conference of VEAMS in CUHK for helpful comments. All errors are ours.

†CUHK (Shenzhen), email: xusichuang@cuhk.edu.cn.
‡CMU, email: shengxiz@andrew.cmu.edu.

1 Introduction

The Human Genome Project (HGP), initiated in 1989 and completed in 2003, provides a natural experiment for understanding how data sharing affects knowledge diffusion. During the project's active phase, genomic sequence data was continuously deposited into public databases, fostering a culture of open data sharing that enabled collaborative breakthroughs across the pharmaceutical industry. Figure 1 shows that the launch of HGP coincided with a marked increase in knowledge diffusion—measured by upward citations from less productive to more productive firms—while the project's completion, after which new public data additions ceased, was followed by a slowdown in diffusion.

Figure 1: Data Sharing and Knowledge Diffusion



Note: This figure shows the time series of upward citations in the pharmaceutical sector. It is measured by firms' upward citation over total backward citation within this sector. The vertical line indicates the year of the launch and the completion of the Human Genome Project.

The HGP experience illustrates a broader tension in modern knowledge economies: while proprietary data and software enhance firms' ability to create new knowledge, they simultaneously create barriers that impede knowledge diffusion to other firms. This paper develops a theoretical framework to understand this double-edged nature of firm-specific intangible capital and its implications for aggregate growth, then applies this framework to study how proprietary data and software affect knowledge diffusion in modern economies.

Theoretical Framework We develop a dynamic general equilibrium model that endogenizes both knowledge creation and diffusion across heterogeneous firms in a quality-ladder environment. The model is tractable because innovation and learning are built from a Poisson-Pareto (extreme-value) structure. In equilibrium, the cross-sectional distribution of idea quality is Fréchet and remains Fréchet after max aggregation, so aggregate dynamics are summarized by a single state variable, the Fréchet scale parameter Λ_t. This delivers a closed-form law of motion for the frontier: innovation and diffusion shift Λ_t through an additive hazard and a moment of firms' knowledge-production productivities.

The same extreme-value structure also yields a hazard-share microfoundation for firm incentives. A firm's probability of contributing the frontier idea has a closed-form expression as its share of aggregate tail intensity, which makes the R&D problem tractable and delivers sharp cross-sectional predictions: high-quality firms tilt toward internal innovation, while low-quality firms tilt toward learning from others.

Endogenous Barriers to Diffusion Section 3 extends the baseline model by allowing firms to invest in data and software that raise the productivity of idea creation but reduce the transferability of ideas across firms. This breaks the baseline additivity between innovation and diffusion hazards and implies that a decline in the cost of firm-specific inputs can increase innovative effort while slowing diffusion and lowering long-run growth.

Application to Data and Software We apply the framework to proprietary data and software. Using U.S. patent citations matched to Compustat firm accounts, we measure vertical knowledge diffusion following Akcigit and Ates (2023) and relate it to leaders' intangible accumulation. Our core firm-level estimate is that a 1% increase in leading firms' software stock reduces upward citations from follower firms by 0.08%, while the response to leaders' R&D is negligible. Section 3 documents additional supporting evidence in time series and sector-level data.

Quantitative Analysis When calibrated to match the empirical patterns we document for data and software, the model generates predictions that align with the observed relationships. We simulate a permanent 30% reduction in the cost of data and software development—

matching the observed decline from 2000 to 2015—and find that the model predicts an immediate but short-lived innovation boom, followed by a long-run decline in aggregate knowledge growth as reduced diffusion dominates enhanced creation. The model also reproduces the differential response of upward citations to R&D versus software investment: increases in leading firms' R&D investment raise upward citations by enhancing knowledge quality, while increases in software investment reduce upward citations by creating stronger diffusion barriers.

The model further distinguishes between two policy-relevant dimensions of data sharing. Consumer privacy protection, which raises the cost of data acquisition, can increase growth by reducing barriers to diffusion. Industrial data sharing policies, which affect firms' willingness to share data with other firms, are overall beneficial for aggregate growth as they facilitate knowledge diffusion across firms. This distinction highlights the different policy implications of consumer privacy regulations versus industrial data sharing policies.

Related Literature and Roadmap Our paper contributes to three strands of literature. First, we build on models of growth with knowledge diffusion across heterogeneous firms (Lucas Jr and Moll, 2014; Perla and Tonetti, 2014; Buera and Oberfield, 2020). Relative to this work, our contribution is a tractable joint theory of creation and diffusion: an extreme-value structure delivers (i) Fréchet preservation and a single-state law of motion for the frontier and (ii) a hazard-share representation of firms' frontier contributions, which pins down innovation incentives without tracking the full cross-sectional state.

A small strand of literature jointly studies firms' incentives to innovate and learn (e.g., König, Lorenz, and Zilibotti (2016) and Hopenhayn, Shi, and Okumura (2025)). As in these papers, we obtain tractable dynamics for the evolution of knowledge. An important difference is that knowledge diffusion in our framework is subject to barriers that are endogenous not only to the level but also to the composition of innovation.

Second, we contribute to work on intangible capital and growth (Corrado et al., 2022; Crouzet et al., 2022; Haskel and Westlake, 2018). We emphasize a channel through which firm-specific intangibles can reduce the transferability of knowledge, weakening diffusion even as they raise the productivity of idea creation.

Third, our paper relates to the literature on the data economy (Jones and Tonetti, 2020; Veldkamp and Chung, 2019). Akcigit and Ates (2021) argue that rising reliance on data and soft-

ware may contribute to declining diffusion; we provide evidence consistent with this mechanism and a quantitative framework that maps observable intangible accumulation into endogenous barriers to learning.

The remainder of the paper is organized as follows. Section 2 develops the baseline model and states the main theoretical results. Section 3 presents the empirical evidence, extends the model to incorporate data and software as endogenous barriers to diffusion, and quantifies the implications for diffusion and growth.

2 A Model of Knowledge Creation and Diffusion

This section develops a dynamic general equilibrium model of knowledge creation and diffusion across heterogeneous firms. The model builds on quality-ladder frameworks in which firms compete to produce the highest-quality version of each variety. Our key innovation is to endogenize both the creation of new knowledge through R&D investment and the diffusion of knowledge across firms, while allowing firm-specific inputs (data and software) to affect both margins. The model generates tractable aggregate dynamics through an extreme-value structure that preserves the Fréchet distribution of idea quality across varieties, enabling closed-form characterizations of equilibrium investment policies and aggregate growth dynamics.

2.1 Economic Environment

Time is discrete and indexed by $t = 0, 1, 2, \ldots$. The economy consists of a representative household, a continuum of varieties indexed by $\omega \in [0, 1]$, a unit mass of inventors, and a continuum of potential producers. Ideas are variety-specific: to produce a given variety, a producer must acquire an idea for that variety. At any point in time, only the producer holding the highest-quality idea for a variety is active. When a new idea improves upon the incumbent quality, competition among potential producers ensures that the inventor extracts the full surplus from selling the idea.

Let $q_t(\omega)$ denote the quality of the best available idea for variety ω at time t. The cross-sectional distribution of qualities across varieties is endogenous and plays a central role in aggregate dynamics. This distribution evolves through two channels: knowledge creation (in-

ternal innovation) and knowledge diffusion (learning from others), both of which we model explicitly below.

Households The representative household has preferences over consumption and labor supply:

$$U = \sum_{t=0}^{\infty} \delta^t \left[u(C_t) - v(H_t) \right], \tag{1}$$

where $\delta \in (0,1)$ is the discount factor, C_t is a composite consumption index, and H_t denotes labor supplied to production of the numeraire good. The consumption aggregator takes the form

$$C_t = \left[\int_0^1 \left(q_t(\omega) c_t(\omega) \right)^{\varepsilon} \, di \right]^{1/\varepsilon}, \qquad \varepsilon \in (0,1), \tag{2}$$

where $q_t(\omega)$ is the quality of variety ω and $c_t(\omega)$ is consumption of that variety. This specification implies that higher-quality varieties contribute more to aggregate consumption, creating market incentives for quality improvement.

The household chooses consumption $\{c_t(\omega)\}$ and labor supply H_t taking prices $\{p_t(\omega), w_t\}$ as given, subject to the budget constraint

$$\int_0^1 p_t(\omega) c_t(\omega) \, d\omega = w_t H_t. \tag{3}$$

Cost minimization yields the standard CES demand schedule:

$$c_t(\omega) = \left(\varepsilon \frac{u'(C_t)}{v'(H_t)} \right)^{\frac{1}{1-\varepsilon}} C_t \, q_t(\omega)^{\frac{\varepsilon}{1-\varepsilon}}, \tag{4}$$

so that demand is increasing in quality. This property ensures that higher-quality varieties command larger market shares, providing the foundation for quality-ladder competition.

Production and Profits Production of each variety uses labor one-for-one. A producer holding an idea of quality $q_t(\omega)$ chooses labor $h_t(\omega)$ to maximize profits:

$$\max_{h_t(\omega)} \; p_t(\omega) h_t(\omega) - w_t h_t(\omega). \tag{5}$$

Optimal pricing implies a constant markup over marginal cost. Under the functional forms we adopt below, profits are a power function of quality:

$$\pi_t(\omega) = (1-\varepsilon)\varepsilon^{\frac{\varepsilon}{1-\varepsilon}} \left(\frac{u'(C_t)}{v'(H_t)}\right)^{\frac{1}{1-\varepsilon}} C_t \, q_t(\omega)^{\frac{\varepsilon}{1-\varepsilon}}. \tag{6}$$

This proportionality is convenient for tractability, as it preserves the monotone mapping between idea quality and private value.

Throughout the paper, we adopt the following functional forms for utility:

$$u(C) = \frac{C^{1-\nu}}{1-\nu}, \qquad v(H) = H, \tag{7}$$

where $\nu > 0$ governs the curvature of consumption utility. These choices imply that equilibrium prices, profits, and the stochastic discount factor admit closed-form expressions; Appendix A.1 collects the demand and profit relationships used in the main text.

Quality Distribution Let $F_t(q)$ denote the cross-sectional distribution of idea quality across varieties at the beginning of period t. Motivated by the extreme-value structure of innovation and diffusion derived below, we conjecture and later verify that the equilibrium distribution belongs to the Fréchet family:

$$F_t(q) = \exp\left(-\Lambda_t q^{-\theta}\right), \qquad \theta > 0. \tag{8}$$

The single state variable Λ_t summarizes the economy's technological frontier and evolves endogenously as a function of firms' innovation and learning decisions. The shape parameter θ governs the tail thickness of the quality distribution: smaller values correspond to fatter tails and greater inequality in idea quality across varieties.

This general equilibrium environment closes the loop between household preferences, firm profits, and the stochastic process governing idea creation and diffusion. Because growth in average quality raises aggregate consumption, it endogenously affects the marginal utility of consumption and therefore the effective discounting of future innovation rents. The equilibrium distribution of idea quality evolves endogenously through the interaction of knowledge creation and diffusion, which we now characterize in detail.

2.2 Knowledge Creation

At the beginning of period t, each firm is characterized by its quality level $q \in \mathbb{R}_+$. Firms can improve their ability to generate new ideas by investing in research and development (R&D), which includes expenditures on goods as well as the development of data and software. The key assumption is that more productive firms are better at transforming R&D resources into high-quality innovations, creating complementarities between existing knowledge and new investment. This assumption captures the empirical regularity that firms with superior existing capabilities tend to generate more valuable innovations from a given level of R&D expenditure.

Let $A_t^k(q)$ denote a firm's productivity in knowledge creation, which depends on both the firm's current quality and its R&D investment. This productivity determines both the rate at which new ideas arrive and the distribution of their quality. The stochastic nature of innovation implies that firms face uncertainty about both the quantity and quality of ideas they generate, reflecting the inherent unpredictability of the innovation process.

R&D Technology A firm of quality q that invests $i \geq 0$ units of the final good in R&D attains knowledge creation productivity:

$$A_t^k(q, i) = q^\alpha i^{1-\alpha}, \qquad 0 < \alpha < 1. \tag{9}$$

This Cobb-Douglas specification captures complementarity between existing firm quality and R&D investment. The parameter α governs the relative importance of existing knowledge versus new investment: when α is large, high-quality firms benefit disproportionately from R&D spending, reflecting stronger complementarities between knowledge and innovation effort.

Let $i_t(q)$ denote the firm's optimal R&D investment policy, which we derive below. For notational simplicity, we define the firm's effective knowledge creation productivity as:

$$A_t(q) \equiv A_t^k(q, i_t(q)). \tag{10}$$

This formulation ensures that more productive firms are more effective at transforming R&D resources into new technological knowledge, generating the cross-sectional heterogeneity in innovation outcomes that we observe in the data.

Idea Generation Knowledge creation is stochastic along two dimensions: the number of ideas a firm generates and the quality of those ideas. This dual uncertainty captures the empirical regularity that innovation outcomes are highly variable, with most ideas having modest impact but a few generating transformative breakthroughs.

Conditional on productivity $A_t(q)$, the number of new ideas generated by the firm follows a Poisson process with intensity:

$$\lambda_t = \kappa_t \bar{z}^{-\theta}, \tag{11}$$

where $\kappa_t > 0$ is a time-varying innovation efficiency parameter and $\bar{z} > 0$ is a scale parameter that we take to zero in the limit to obtain a continuous distribution of idea quality.

Each idea's quality is given by:

$$q^{\text{new}} = z A_t(q)^{\rho}, \qquad 0 \leq \rho \leq 1, \tag{12}$$

where z is an i.i.d. random variable drawn from a Pareto distribution with cumulative distribution function:

$$\Pr(z \leq x) = 1 - \left(\frac{x}{\bar{z}}\right)^{-\theta}, \qquad x \geq \bar{z}, \tag{13}$$

for tail parameter $\theta > 0$. Higher knowledge creation productivity shifts the quality distribution of new ideas upward in the sense of first-order stochastic dominance. The parameter ρ governs the extent to which productivity improvements translate into higher-quality ideas: when $\rho = 1$, the relationship is proportional, while smaller values of ρ imply diminishing returns to productivity in idea quality.

Distribution of New Ideas Let $H_{kt}^{\text{new}}(q \mid A_t(\widetilde{q}))$ denote the distribution of the quality of a single new idea generated by a firm with productivity $A_t(\widetilde{q})$. This distribution satisfies

$$\begin{aligned} H_{kt}^{\text{new}}(q \mid A_t(\widetilde{q})) &= \Pr\left(z \leq \frac{q}{A_t(\widetilde{q})^{\rho}}\right) \\ &= \left[1 - \left(\frac{q}{\bar{z} A_t(\widetilde{q})^{\rho}}\right)^{-\theta}\right] \mathbb{I}\left(A_t(\widetilde{q}) \leq \left(\frac{q}{\bar{z}}\right)^{1/\rho}\right). \end{aligned} \tag{14}$$

Best Idea Created Let $H_{kt}^{\text{best}}(q \mid A_t(\widetilde{q}))$ denote the distribution of the best idea among all ideas created by a firm with productivity $A_t(\widetilde{q})$. Since the number of ideas is Poisson distributed,

standard extreme-value arguments imply

$$H_{kt}^{\text{best}}(q \mid A_t(\widetilde{q})) = \sum_{s=0}^{\infty} \frac{\lambda_t^s e^{-\lambda_t}}{s!} \left[H_{kt}^{\text{new}}(q \mid A_t(\widetilde{q}))\right]^s$$
$$= \exp\{-\lambda_t \left[1 - H_{kt}^{\text{new}}(q \mid A_t(\widetilde{q}))\right]\}. \tag{15}$$

Substituting for λ_t yields

$$H_{kt}^{\text{best}}(q \mid A_t(\widetilde{q})) = \exp\left\{-\kappa_t \left(\frac{q}{A_t(\widetilde{q})^\rho}\right)^{-\theta} \mathbb{I}\left(A_t(\widetilde{q}) \leq \left(\frac{q}{\bar{z}}\right)^{1/\rho}\right)\right\}. \tag{16}$$

Aggregation and Limiting Economy Let F_t denote the cross-sectional distribution of firm quality at time t. The unconditional distribution of a new idea is given by

$$H_{kt}^{\text{new}}(q) = \int H_{kt}^{\text{new}}(q \mid A_t(\widetilde{q}))\, dF_t(\widetilde{q}). \tag{17}$$

We consider the limit as $\bar{z} \to 0$, in which individual ideas become arbitrarily small but arrive at an arbitrarily high rate. This limiting procedure, standard in extreme-value theory, ensures that the aggregate distribution of the best idea remains well-defined and follows a Fréchet distribution. In this limit,

$$\kappa_t \bar{z}^{-\theta} \left[1 - H_{kt}^{\text{new}}(q)\right] \longrightarrow \kappa_t \int_0^{\infty} \left(\frac{q}{A_t(\widetilde{q})^\rho}\right)^{-\theta} dF_t(\widetilde{q}). \tag{18}$$

Consequently, the limiting distribution of the best idea created in the economy satisfies

$$H_{kt}^{\text{best}}(q) = \exp\left\{-\kappa_t \int_0^{\infty} \left(\frac{q}{A_t(\widetilde{q})^\rho}\right)^{-\theta} dF_t(\widetilde{q})\right\}. \tag{19}$$

This expression shows that the aggregate distribution depends on the cross-sectional distribution of firm productivities, weighted by their contribution to knowledge creation.

Economic Interpretation This formulation embeds endogenous innovation incentives into a quality-ladder environment in the spirit of Akcigit, Celik, and Greenwood (2016) and related work. Firms invest in R&D to increase their arrival rate of high-quality ideas, where the effec-

tiveness of R&D is complementary to firm quality. Innovation is characterized by a fat-tailed distribution of outcomes, so that growth is driven by rare but transformative ideas. Aggregate technological progress emerges from the interaction between the cross-sectional distribution of firm productivities and firms' R&D decisions, generating a tractable hazard-rate representation of creative activity that will later interact with knowledge diffusion. The Poisson-Pareto structure ensures that the distribution of the best idea across all firms remains tractable while capturing the right-skewed nature of innovation outcomes.

2.3 Knowledge Diffusion

Firms can improve their knowledge not only through internal R&D but also by learning from other firms. We model diffusion using a Poisson meeting structure. In the baseline model, innovation and diffusion share the same extreme-value structure and differ only in their hazard rates (κ_t and μ_t), which preserves tractability in the law of motion for the frontier. Section 3 introduces barriers that make ideas less transferable and break this symmetry.

Meetings and the External Idea Draw In each period t, firms encounter ideas from other firms according to a Poisson process. Let S_t denote the number of distinct external ideas that a firm encounters in period t. We assume:

$$S_t \sim \text{Poisson}(\mu_t \, \bar{z}^{-\theta}), \tag{20}$$

where $\mu_t > 0$ is the diffusion intensity and $\bar{z}$ is the upper scale parameter of the underlying idea draw.

When a firm meets a source firm with beginning-of-period product quality $\widetilde{q}$, it observes an idea draw whose quality is scaled by the source's effective knowledge-creation productivity. Let $i_t(q)$ denote the R&D investment policy function, and let $A_t(q)$ denote the internal knowledge-production productivity (as defined in Section 2.2). Conditional on meeting a source firm of type $\widetilde{q}$, the quality of a *new* external idea satisfies:

$$q^{\text{new}} \;=\; z \, A_t(\widetilde{q})^{\rho}, \tag{21}$$

where z is drawn from a Pareto/Fréchet-type distribution with tail parameter $\theta > 0$. This specification ensures that more productive source firms generate higher-quality ideas that are available for diffusion. Equivalently, the conditional cumulative distribution function of a *single* external idea draw is:

$$\begin{aligned} H_{dt}^{\text{new}}(q \mid A_t(\widetilde{q})) &= \Pr\left(z \leq \frac{q}{A_t(\widetilde{q})^{\rho}}\right) \\ &= \left[1 - \left(\frac{q}{\bar{z}\, A_t(\widetilde{q})^{\rho}}\right)^{-\theta}\right] \mathbb{I}\left(A_t(\widetilde{q}) \leq \left(\frac{q}{\bar{z}}\right)^{1/\rho}\right). \end{aligned} \tag{22}$$

Let F_t denote the cross-sectional distribution of beginning-of-period product quality. The unconditional distribution of a new external idea draw is then the mixture over sources,

$$H_{dt}^{\text{new}}(q) = \int_0^{A_t^{-1}\left[(q/\bar{z})^{1/\rho}\right]} H_{dt}^{\text{new}}(q \mid A_t(\widetilde{q}))\; dF_t(\widetilde{q}). \tag{23}$$

Best External Idea and Poisson Aggregation Given $S_t = s$ meetings, the best external idea is the maximum of s i.i.d. draws from $H_{dt}^{\text{new}}(q)$, hence has CDF $H_{dt}^{\text{new}}(q)^s$. Averaging over the Poisson distribution delivers the standard hazard-form expression:

$$\begin{aligned} H_{dt}^{\text{best}}(q) &= \sum_{s=0}^{\infty} \Pr(S_t = s)\, H_{dt}^{\text{new}}(q)^s = \sum_{s=0}^{\infty} \frac{(\mu_t \bar{z}^{-\theta})^s e^{-\mu_t \bar{z}^{-\theta}}}{s!}\, H_{dt}^{\text{new}}(q)^s \\ &= \exp\left\{-\mu_t \bar{z}^{-\theta}\left[1 - H_{dt}^{\text{new}}(q)\right]\right\}. \end{aligned} \tag{24}$$

The *internal* knowledge-creation channel is analogous, with Poisson intensity $\kappa_t \bar{z}^{-\theta}$ and the same idea-quality distributional form (the distinction between creation and diffusion is purely in the hazard rates):

$$H_{kt}^{\text{best}}(q) = \exp\left\{-\kappa_t \bar{z}^{-\theta}\left[1 - H_{kt}^{\text{new}}(q)\right]\right\}, \qquad H_{kt}^{\text{new}}(q) = H_{dt}^{\text{new}}(q). \tag{25}$$

Implication for the Law of Motion of Product Quality The following proposition summarizes the tractability result that underpins aggregate dynamics in the baseline model.

Proposition 1 (Fréchet preservation and a single-state law of motion)**.** *Suppose the beginning-of-period distribution of product quality is Fréchet,* $F_t(q) = \exp\left(-\Lambda_t q^{-\theta}\right)$*. Under max aggrega-*

tion across the incumbent and the best ideas drawn from knowledge creation and diffusion, the Fréchet family is preserved. In particular, the best new *idea is Fréchet with scale increment* $\Delta\Lambda_t$ *in* (28)*, and the next-period distribution satisfies* (29)*.*

Suppose that the beginning-of-period distribution of product quality is Fréchet:

$$F_t(q) = \exp\left(-\Lambda_t q^{-\theta}\right). \tag{26}$$

Under max-aggregation across (i) the incumbent quality and (ii) the best ideas drawn from knowledge creation and diffusion, the Fréchet family is preserved. In particular, the distribution of the best *new* idea available to a firm in period t is

$$H_{dt}^{\text{best}}(q)\, H_{kt}^{\text{best}}(q) = \exp\left\{-(\kappa_t + \mu_t)\int_0^\infty \left(\frac{q}{A_t(\widetilde{q})^\rho}\right)^{-\theta} dF_t(\widetilde{q})\right\} = \exp\left(-\Delta\Lambda_t\, q^{-\theta}\right), \tag{27}$$

where the increment to the Fréchet scale parameter is

$$\Delta\Lambda_t = (\kappa_t + \mu_t)\int_0^\infty A_t(\widetilde{q})^{\rho\theta}\, dF_t(\widetilde{q};\Lambda_t). \tag{28}$$

Consequently, the next-period distribution remains Fréchet with updated parameter

$$F_{t+1}(q) = \exp\left(-\Lambda_{t+1} q^{-\theta}\right), \qquad \Lambda_{t+1} = \Lambda_t + \Delta\Lambda_t. \tag{29}$$

Economic Interpretation Knowledge diffusion is disciplined by a symmetric meeting technology: all firms draw from the same distribution of external idea quality, but they differ in the number of opportunities they have to learn. In the baseline model (where data&software choices do not directly distort the transmissibility of ideas), the framework cleanly separates the *rate* of internal innovation (κ_t) from the *rate* of learning from others (μ_t). This separation is crucial for understanding how different policies or technological changes affect knowledge accumulation through these distinct channels.

Equation (28) shows that the two hazards enter additively, so innovation and diffusion jointly shift the scale parameter of the Fréchet distribution while leaving its shape parameter θ unchanged. This additive structure implies that innovation and diffusion are perfect substitutes

at the margin in terms of their contribution to aggregate technological progress, though they may have different implications for the cross-sectional distribution of firm quality. The extension in Section 3 introduces barriers to diffusion that break this perfect substitutability, allowing us to study how firm-specific investments in data and software create asymmetric effects on knowledge creation versus knowledge diffusion.

2.4 Firm R&D Incentives and Frontier Contributions

This section formulates the innovating firm's dynamic optimization problem and then microfounds the key object $k_t(q, i)$—the probability (or, equivalently in the Poisson-race formulation, the expected number) of frontier contributions generated by a firm of beginning-of-period quality q after choosing R&D investment i. This object captures how a firm's investment decisions affect its probability of generating the economy-wide frontier idea, which determines the firm's continuation value. The microfoundation relies on a hazard-share argument that exploits the extreme-value structure of the innovation process. This approach provides a tractable characterization of investment incentives without requiring explicit computation of the full distribution of candidate ideas across all firms.

Firm Problem Let $W_t(q)$ denote the value of a firm with state q at the beginning of period t. In period t, the firm chooses R&D investment $i \geq 0$. With probability $k_t(q, i)$, the firm contributes the frontier idea in that period; conditional on contributing, the quality of the frontier idea is distributed according to the economy-wide frontier distribution $H_t^{best}(\cdot)$ derived below.

When the firm contributes a frontier idea of quality $\widetilde{q}$, it becomes the new incumbent with quality $\widetilde{q}$ next period. If the firm does not contribute, it disappears. If it contributes multiple frontier ideas, it splits into multiple firms, each managing a single frontier idea. Let M_t denote the stochastic discount factor. The Bellman equation is:

$$W_t(q) = \max_{i \geq 0} \Big\{ -i + M_t \, k_t(q, i) \int_0^\infty \left[\pi_{t+1}(\widetilde{q}) + W_{t+1}(\widetilde{q}) \right] \, dH_t^{best}(\widetilde{q}) \Big\}. \tag{30}$$

The endogenous margin is $k_t(q, i)$, which captures how current investments tilt the firm's odds in the frontier race. The firm's optimization problem thus trades off the cost of R&D

investment against the increase in the probability of capturing the frontier, weighted by the expected continuation value.

Microfoundation of $k_t(q,i)$ via a Poisson Race We now derive $k_t(q,i)$ from primitives. The argument proceeds in three steps: (i) represent candidate ideas as a Poisson point process over quality, (ii) characterize the economy-wide frontier distribution, and (iii) use a hazard-share identity to determine which source generates the maximum. This approach exploits the extreme-value structure of the innovation process to obtain a tractable closed-form expression for the probability of frontier contribution.

Step 1: Candidate ideas as Poisson point processes. Fix period t and a quality threshold $x \geq 1$. Candidate ideas come from two sources: internal knowledge creation and external knowledge diffusion. We characterize each source as a Poisson point process over the quality space.

(i) Knowledge creation. A firm of type q choosing i produces internal candidate ideas according to a Poisson process with intensity $\kappa_t > 0$. The right tail of a single candidate draw is Fréchet/Pareto in the sense that

$$\Pr(X > x \mid q,i) = A_t^{\kappa}(q,i)^{\rho\theta}\, x^{-\theta}, \qquad \theta > 0. \tag{31}$$

Hence the *number* of internal candidate ideas exceeding x is Poisson with mean

$$\lambda_t^{\kappa}(q,i;x) = \kappa_t\, A_t^{\kappa}(q,i)^{\rho\theta}\, x^{-\theta}. \tag{32}$$

(ii) Knowledge diffusion. Independently, the firm receives learned candidate ideas from diffusion. Aggregating over the cross-section, let the diffusion arrival intensity be $\mu_t > 0$ and let the effective tail scale of learned candidates be

$$\overline{\mu}_t \equiv \int_0^{\infty} A_t^{\mu}(\widetilde{q})^{\rho\theta}\, dF_t(\widetilde{q}). \tag{33}$$

Then the number of learned candidates exceeding x is Poisson with mean

$$\lambda_t^{\mu}(x) = \mu_t\, \overline{\mu}_t\, x^{-\theta}. \tag{34}$$

(iii) Economy-wide creation scale. Define the cross-sectional average creation tail scale

$$\overline{\kappa}_t \equiv \int_0^\infty A_t^\kappa(\widetilde{q}, i_t(\widetilde{q}))^{\rho\theta}\, dF_t(\widetilde{q}). \tag{35}$$

Then the *total* expected number of (economy-wide) internal candidates exceeding x is $\kappa_t\overline{\kappa}_t x^{-\theta}$, and the total expected number of learned candidates exceeding x is $\mu_t\overline{\mu}_t x^{-\theta}$.

Step 2: The frontier distribution. Let $Q_t^{\max}$ denote the maximum quality among *all* candidate ideas in period t. The event $\{Q_t^{\max} \le x\}$ is equivalent to *no* candidate idea exceeding x. Since counts of exceedances from independent Poisson sources add, the total number of candidates exceeding x is Poisson with mean

$$x^{-\theta}\big(\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t\big).$$

Therefore,

$$H_t^{best}(x) \equiv \Pr(Q_t^{\max} \le x) = \exp\Big\{ - x^{-\theta}\big(\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t\big)\Big\}. \tag{36}$$

Differentiating yields the frontier density

$$h_t^{best}(x) = \theta\big(\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t\big)x^{-(1+\theta)} \exp\Big\{ - x^{-\theta}\big(\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t\big)\Big\}. \tag{37}$$

Step 3: Who produces the maximum—an explicit hazard-share calculation. We now compute the probability that the firm of type q is the source of the maximum, which is the object $k_t(q,i)$ in (30). The clean way to see the hazard-share logic is to work with an infinitesimal quality interval.

Fix x and consider the event that the frontier falls in $(x, x+dx)$. This requires two things:

1. No candidate idea exceeds $x+dx$ (so the maximum is below $x+dx$), and
2. At least one candidate idea falls in $(x, x+dx)$.

For Poisson point processes, the probability of an event in an infinitesimal interval is governed by the *intensity density* at x. From (32)–(34), the expected number of ideas in $(x, x+dx)$ coming from a source with tail scale S is

$$d\lambda(x) = \theta S\, x^{-(1+\theta)}\, dx.$$

Hence:

- The firm's *own* intensity density at x equals

$$d\lambda_t^{firm}(x) = \theta\big(\kappa_t A_t^{\kappa}(q,i)^{\rho\theta} + \mu_t\overline{\mu}_t\big)x^{-(1+\theta)}dx.$$

- The *economy-wide* intensity density at x equals

$$d\lambda_t^{total}(x) = \theta\big(\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t\big)x^{-(1+\theta)}dx.$$

Now use independence. Conditional on the event that *some* idea lands in $(x, x+dx)$ and that nothing exceeds $x+dx$, the probability that the idea in $(x, x+dx)$ comes from a given source is the ratio of that source's intensity density to the total intensity density:

$$\Pr(\text{winner is firm} \mid Q_t^{\max} \in (x, x+dx)) = \frac{d\lambda_t^{firm}(x)}{d\lambda_t^{total}(x)} = \frac{\kappa_t A_t^{\kappa}(q,i)^{\rho\theta} + \mu_t\overline{\mu}_t}{\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t}. \tag{38}$$

Crucially, because all sources share the same power tail $x^{-(1+\theta)}$, the ratio is *independent of* x. Therefore the firm's probability of supplying the frontier idea is constant across frontier levels and equals the hazard share in (38). This is the desired object.

Proposition 2 (Hazard-share representation of frontier contributions)**.** *In period t, in the baseline model the probability that a firm of type q supplies the frontier idea equals its share of aggregate tail intensity:*

$$k_t(q,i) = \frac{\kappa_t A_t^{\kappa}(q,i)^{\rho\theta} + \mu_t\overline{\mu}_t}{\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t}. \tag{39}$$

Interpretation and Link Back to Incentives. Equation (39) shows that R&D investments affect the firm's dynamic problem only by raising $A_t^{\kappa}(q,i)$—the scale of the firm's internal idea tail. Because the frontier is an extreme-value object, the marginal return to investment i is a *race* effect: higher investment raises the firm's share of the economy-wide tail intensity and thus increases the probability of capturing the frontier in that period. This hazard-share interpretation is central to the model's tractability: it allows us to characterize the firm's investment incentives without explicitly solving for the distribution of all candidate ideas. Substituting (39) into the Bellman equation (30) yields a standard quality-ladder incentive condition in which

the marginal cost of investment is equated to the marginal increase in the probability of moving to the frontier times the continuation value gain. The key insight is that firms compete in a race to generate the highest-quality idea, and investment increases their probability of winning this race in proportion to their share of the aggregate innovation intensity.

2.5 The Dynamic Equilibrium

We now characterize the full dynamic equilibrium of the economy, followed by a discussion of the underlying mechanism as well as the time-series and cross-sectional implications.

Characterization of the Equilibrium The equilibrium consists of sequences of value functions, quality distributions, and investment policies that satisfy the firm's optimization problem and the law of motion for the aggregate state variable Λ_t. The key insight is that the equilibrium can be decomposed into trend and level components, where the trend captures balanced growth and the level captures deviations from trend. This decomposition is essential for analytical tractability: it allows us to separate the long-run growth path from cyclical fluctuations, while maintaining a unified framework that endogenously determines both.

Proposition 3. *Assume processes for κ_t and μ_t take the form $\kappa_t = \widehat{\kappa}_t g^t$ and $\mu_t = \widehat{\mu}_t g^t$, where $\widehat{\kappa}_0$ and $\widehat{\mu}_0$ are given and $\widehat{\kappa}_t$ and $\widehat{\mu}_t$ are stationary processes. Given initial condition Λ_0, the equilibrium system consists of sequences $\left\{\Lambda_t^*, W_t^*, \widehat{\Lambda}_t, \widehat{W}_t, \Lambda_t, W_t\right\}_{t=0}^{\infty}$ and W_0 that satisfy the decomposition $\Lambda_t = \Lambda_t^* \widehat{\Lambda}_t$ and $W_t = W_t^* \widehat{W}_t$, where the starred variables are the trend components (under the condition that $1 - \alpha\rho - \frac{1-\nu}{\nu}(1-\alpha)\rho > 0$)*

$$\Lambda_t^* = g_\Lambda^t,\ W_t^* = g_\Lambda^{\frac{1}{\theta}\frac{1-\nu}{\nu}t}, \tag{40}$$

where $g_\Lambda \equiv g^{\frac{1}{1-\alpha\rho-\frac{1-\nu}{\nu}(1-\alpha)\rho}}$*, and the hatted variables are the level component,*

$$\widehat{W}_t = \delta\left(g_\Lambda \frac{\widehat{\Lambda}_{t+1}}{\widehat{\Lambda}_t}\right)^{\frac{1}{\theta}\frac{1-2\nu}{\nu}}\left[1-\frac{(1-\alpha)\rho\theta}{1+\widetilde{\mu}_t}\right]\left[(1-\varepsilon)\varepsilon^{\frac{1-\nu}{\nu}}\Gamma\left(1-\frac{1}{\theta}\frac{\varepsilon}{1-\varepsilon}\right)^{\frac{1-\varepsilon}{\varepsilon}\frac{1-\nu}{\nu}}+\widehat{W}_{t+1}\right], \tag{41}$$

$$\widehat{\Lambda}_{t+1} = \frac{[\rho\theta(1-\alpha)]^{(1-\alpha)\rho\theta}\, g_\Lambda^{-1}(1+\widetilde{\mu}_t)}{([1-(1-\alpha)\rho\theta]+\widetilde{\mu}_t)^{(1-\alpha)\rho\theta}}\widehat{\kappa}_t \widehat{W}_t^{\rho\theta(1-\alpha)}\widehat{\Lambda}_t^{\alpha\rho+\frac{1-\nu}{\nu}\rho(1-\alpha)}\left[\Gamma\left(1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right)\right]^{1-\rho\theta(1-\alpha)} \tag{42}$$

where $\widetilde{\mu}_t \equiv \mu_t\overline{\mu}_t/(\kappa_t\overline{\kappa}_t)$ *(which reduces to* μ_t/κ_t *in the baseline model without diffusion barriers) and* $\widehat{W}_0$ *are determined such that* $\lim_{t\to\infty}\widehat{W}_t = \widehat{W}_\infty$*, plus a transversality condition* $\delta < g^{\frac{1}{1-\alpha\rho-\frac{1-\nu}{\nu}(1-\alpha)\rho}\frac{1}{\theta}\frac{2\nu-1}{\nu}}$.

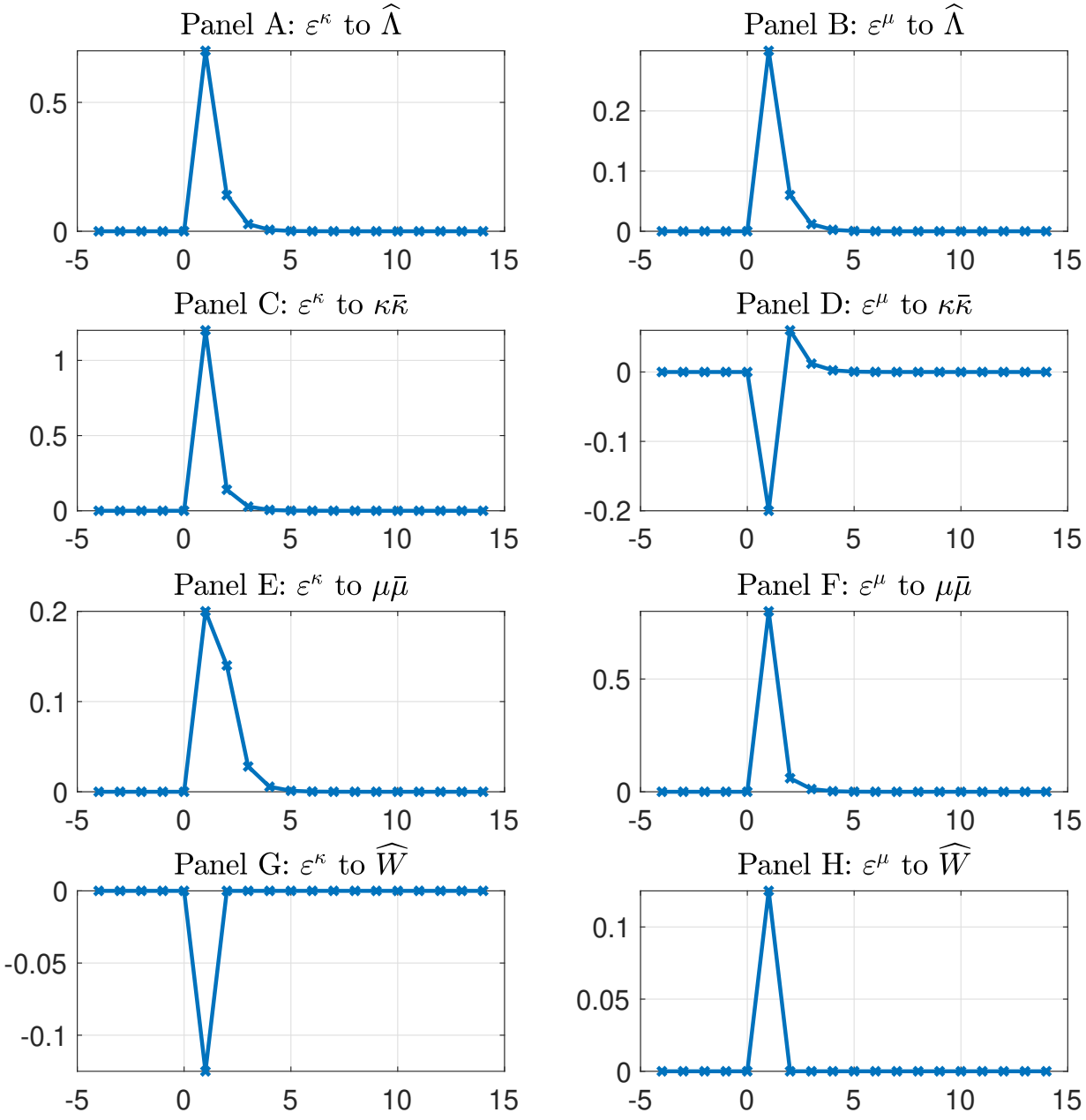
Proof. See Appendix A.3. ■

Inspecting the Mechanism To illustrate the model's mechanism, we consider a simple specification where the detrended shocks follow $\widehat{\kappa}_t = \kappa e^{\varepsilon_t^\kappa}$ and $\widehat{\mu}_t = \mu e^{\varepsilon_t^\mu}$, where ε_t^κ and ε_t^μ are independent i.i.d. shocks that follow normal distributions with variances σ^κ and σ^μ, respectively. With this specification, we can log-linearize the system to obtain an explicit form for the dynamics of $\widehat{\Lambda}_t$ by examining its log deviations from steady state. For tractability, we assume that $\widehat{W}_{t+1}$ reverts to steady state immediately after the shock, which is a standard approximation in linearized dynamic models that captures the essential dynamics while maintaining analytical tractability. This simplification allows us to derive closed-form expressions for impulse response functions that illuminate how innovation and diffusion shocks propagate through the economy.

Corollary 1. *Let* $\widehat{\lambda}_t \equiv \log\widehat{\Lambda}_t - \log\widehat{\Lambda}_\infty$ *represents the log-deviation of knowledge stock from its long-run steady state, then*

$$\widehat{\lambda}_{t+1} = \frac{\rho}{1-\rho(1-\alpha)\frac{1-2\nu}{\nu}}\widehat{\lambda}_t + \frac{1}{\kappa+\mu}\left[\frac{1-(1-\alpha)\rho\theta}{1-\rho(1-\alpha)\frac{1-2\nu}{\nu}}\mu\varepsilon_t^\mu + \frac{\kappa+\mu(1-\alpha)\rho\theta}{1-\rho(1-\alpha)\frac{1-2\nu}{\nu}}\varepsilon_t^\kappa\right]. \tag{43}$$

Figure 2: Impulse and Response Functions



Note: This figure shows the impulse response function of the two shocks. The size of the shock is unit standard deviation. The parameters used to plot this figure are $\nu = 0.50$, $\alpha = 0.50$, $\beta = 0.40$, $\theta = 4.00$, $\rho = 0.20$, $\delta = 0.90$, $\varepsilon = 0.50$, $\kappa = 1.00$, $\mu = 5.00$, $\sigma^{\kappa} = \sigma^{\mu} = 1.0$, and $g = 1.02$.

Figure 2 shows the impulse response functions. The first panel displays the IRFs of $\widehat{\lambda}_t$ to a one standard deviation shock to innovation κ_t and diffusion μ_t. Following the result in equation (43), we can see that both leads to a persistent increase in $\widehat{\Lambda}_t$.

Panel C to Panel F show the innovation and diffusion components behind the dynamics of $\widehat{\lambda}_t$. The contrast between Panel E and D is interesting: although positive shocks to innovation benefits diffusion, positive shocks to diffusion induces a temporal drop of innovation. To understand this temporal drop, we write down the aggregate innovation intensity $\overline{\kappa}_t$:

$$\overline{\kappa}_t = \left(\frac{\rho\theta\left(1-\alpha\right)\kappa_t \overline{W}_t}{\left[1-\left(1-\alpha\right)\rho\theta\right]\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t} \right)^{\frac{\rho\theta(1-\alpha)}{1-(1-\alpha)\rho\theta}} \int_0^\infty \widetilde{q}^{\frac{\alpha\rho\theta}{1-(1-\alpha)\rho\theta}} dF_t\left(\widetilde{q}\right)$$

which is driven by two forces on impact (i) shocks κ_t and μ_t, (ii) the value function $\overline{W}_t$. When μ_t jumps, the firm value increases, so incentive to R&D increases. This is captured by the fact that $\overline{W}_t$ increase on impact. However, when knowledge diffuse faster, the equilibrium knowledge frontier also improves, making a firm's ideas arrive on the frontier less likely. Under certain conditions, we show the latter force dominates, so that there is a transitory drop in innovation upon diffusion increase. The following corollary summarizes this interaction between innovation and diffusion.

Corollary 2. *While permanent increases of diffusion* (μ_t) *always lead to increases in innovation* ($\bar{\kappa}$) *in the long run, a transitory jump can reduce innovation on impact.*

Innovation and Diffusion at the Cross Section Our model contributes to a small but influential strand of literature on cross-sectional innovation and diffusion/imitation incentives (e.g., König, Lorenz, and Zilibotti (2016)). Similar to these models, our framework implies that productive innovators rely more on innovation, whereas smaller innovators rely more on diffusion. We summarize the cross-sectional implications as follows.

For an inventor with quality q, the incentive to innovate (scaled by quality) is given by:

$$\frac{i}{q} = \left\{ \frac{\rho\theta\left(1-\alpha\right)\kappa_t \overline{W}_t}{\left[1-\left(1-\alpha\right)\rho\theta\right]\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t} \right\}^{\frac{1}{1-(1-\alpha)\rho\theta}} q^{\frac{\rho\theta-1}{1-(1-\alpha)\rho\theta}}.$$

Note that i is an increasing function of q. Further, if $1 < \rho\theta < 1/\left(1-\alpha\right)$, then i/q is also increasing with q, implying that high-quality inventors have larger incentives to innovate relative to their quality level. This pattern reflects the complementarity between existing knowledge and R&D investment: firms with superior knowledge stocks benefit more from additional innovation effort.

To what extent do firms rely on diffusion? For a inventor with quality q, the model implies that the likelihood of citing others' ideas is:

$$\frac{\mu_t \overline{\mu}_t}{\kappa_t \left(q^\alpha i^{1-\alpha}\right)^{\rho\theta} + \mu_t \overline{\mu}_t},$$

where $\overline{\mu}_t$ is the aggregate diffusion scale. This expression is a decreasing function of q, indicating that productive innovators rely less on imitation and more on innovation. This prediction is consistent with empirical evidence on the relationship between firm size and innovation strategies, where larger firms tend to engage in more internal R&D relative to learning from others. The model thus generates a natural sorting: high-quality firms invest intensively in internal innovation, while low-quality firms rely more heavily on knowledge diffusion to catch up.

2.6 Additional Discussions

The Case of Exogenous Innovation The model of Buera and Oberfield (2020) can be viewed as a special case of our framework with exogenous innovation. Specifically, if we set $i = \iota q$, where ι is a constant, then the Λ dynamics will be isomorphic to Buera and Oberfield (2020). This specification has two important implications. First, there are no cross-sectional variations in innovation incentives, as all firms invest proportionally to their quality. Second, this setting rules out how aggregate knowledge changes in response to innovation incentives, as investment is mechanically tied to quality rather than being chosen optimally in response to market conditions. Our endogenous specification allows for richer dynamics and cross-sectional heterogeneity, enabling the model to capture how changes in the economic environment (such as declining costs of data and software) affect both the level and distribution of innovation effort across firms.

The Case of $1 - \alpha\rho - \frac{1-\nu}{\nu}(1-\alpha)\rho = 0$ In this special case, the model exhibits endogenous growth driven entirely by endogenous innovation, even in the absence of exogenous technological progress. To demonstrate this point, let $g = 1$ so that there is no exogenous growth in κ_t nor

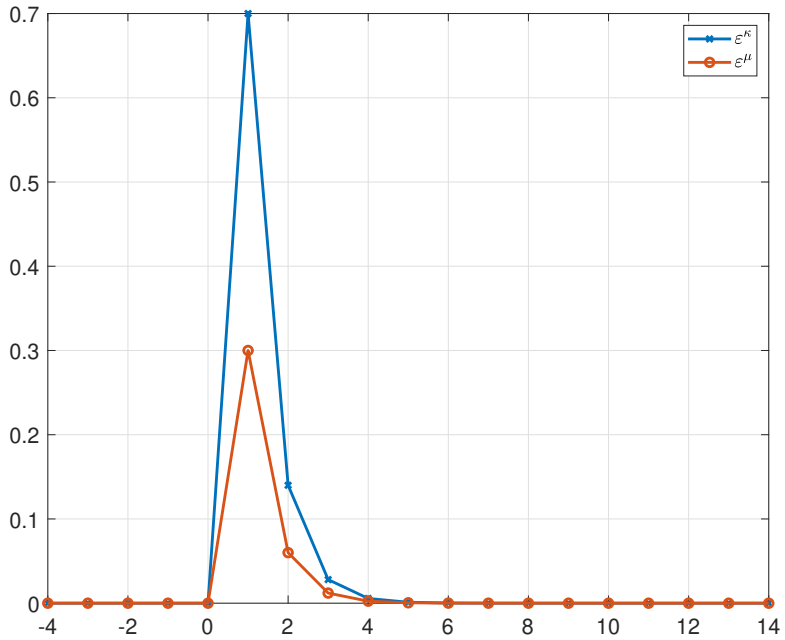
μ_t. We show that the growth rate of Λ_t, denoted $g_{\Lambda t} = \Lambda_{t+1}/\Lambda_t$, solves the following system:

$$\widehat{W}_t = \delta g_{\Lambda t}^{\frac{1}{\theta}\frac{1-2\nu}{\nu}} \left(1 - \frac{\rho\theta\left(1-\alpha\right)}{1+\widetilde{\mu}_t}\right)\left[\left(1-\varepsilon\right)\varepsilon^{\frac{1-\nu}{\nu}}\Gamma\left(1-\frac{\varepsilon}{\theta\left(1-\varepsilon\right)}\right)^{\frac{1-\varepsilon}{\varepsilon}\frac{1-\nu}{\nu}} + \widehat{W}_{t+1}\right],$$

$$g_{\Lambda t} = \left(1+\widetilde{\mu}_t\right)\kappa\left\{\frac{\left(1-\alpha\right)\rho\theta\widehat{W}_t}{\left[1-\left(1-\alpha\right)\rho\theta\right]+\widetilde{\mu}_t}\right\}^{(1-\alpha)\rho\theta}\left[\Gamma\left(1-\frac{\alpha\rho}{1-\rho\theta\left(1-\alpha\right)}\right)\right]^{1-\rho\theta(1-\alpha)}.$$

where $\widehat{W}_t$ is the normalized value function. The steady state value can be computed from the above system. Figure 3 shows the impulse response function. We observe that i.i.d. shocks can lead to persistent declines in the *growth rate*, illustrating how temporary disturbances can have long-lasting effects on economic growth through their impact on innovation incentives. This mechanism provides a microeconomic foundation for growth slowdowns that persist beyond the duration of the initial shock, as the endogenous response of innovation effort amplifies and propagates the initial disturbance through the knowledge accumulation process.

Figure 3: Impulse Response Function to the Growth of Λ_t



Note: This figure shows the impulse response function of the growth rate of knowledge Λ_t to a one time shock. The parameters use are the same as Figure 2, except that we set $g = 1$ and $1 - \alpha\rho - \frac{1-\nu}{\nu}\left(1-\alpha\right)\rho = 0$.

Extensive Margins The model can be extended to include extensive margins by introducing fixed costs to R&D investment. We can show that with extensive margins, some low-productivity firms have zero innovation, while the model remains tractable. This extension allows the model to capture the empirical fact that not all firms engage in R&D, and it generates additional cross-sectional heterogeneity in innovation behavior beyond what is driven by quality differences alone.

3 Application: Data and Software as Endogenous Barriers to Diffusion

This section extends the baseline model to incorporate data and software as endogenous inputs that simultaneously facilitate knowledge creation and impede knowledge diffusion. We document stylized facts linking software accumulation by leading firms to weaker diffusion, then extend the model to include data&software investment and an endogenous barrier to learning, and finally quantify the implications for diffusion and growth.

3.1 Stylized Facts

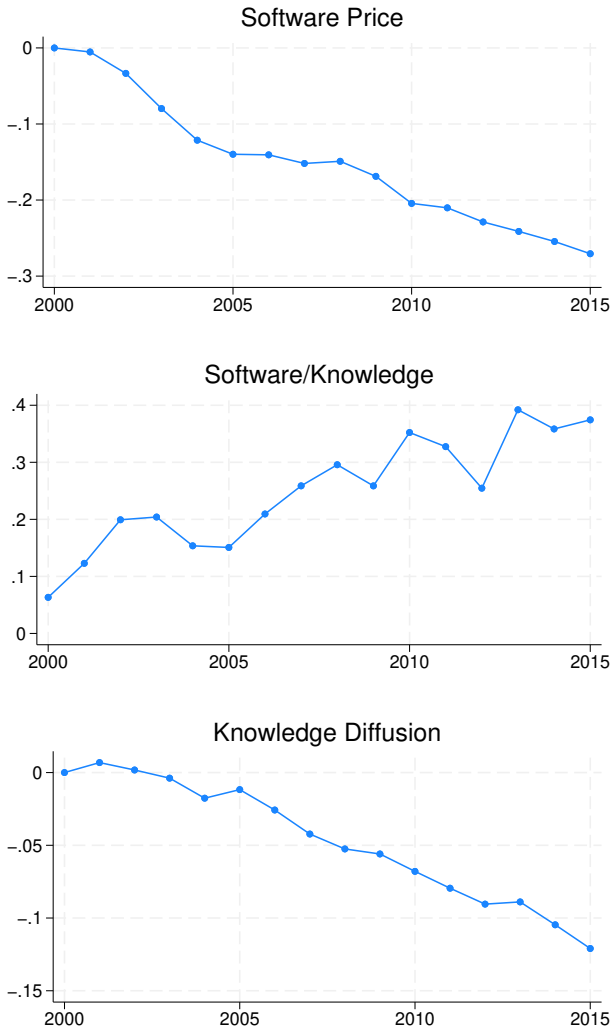
We document three stylized facts that motivate our extension of the baseline model. These facts establish a robust negative relationship between software investment and knowledge diffusion across multiple levels of aggregation. Our empirical analysis merges patent-level data from PatentsView with firm-level accounting data from Compustat. We follow Dorn et al. (2020) to assign patents to firms by matching standardized company names through the Bing search API. Following Akcigit and Ates (2023), we construct measures of knowledge diffusion by examining citation flows from leading firms (those with higher knowledge capital) to follower firms. At the firm level, our key variables are knowledge capital, measured as capitalized R&D expenditure following Peters and Taylor (2007), and software stock, measured as capitalized software expenditure. We now turn to the empirical patterns.

Secular Trends Figure 4 presents three time series that characterize the evolution of software investment and knowledge diffusion. The first panel shows the logarithm of software prices

from the Bureau of Economic Analysis (BEA), which measures the cost of developing data and software using the Hall-Jorgenson user cost model (Garner et al., 2021). From 2000 to 2015, software prices declined by more than 30%, reflecting technological progress in software development. The second panel displays firm-level software intensity, defined as the ratio of capitalized software expenditure to capitalized R&D expenditure, averaged across all firms in the Compustat dataset. Software intensity increased substantially over this period, rising from approximately 3% in 2000 to 15% in 2015.

Fact 1. *Since 2000, the decrease in software price has been accompanied by increased software intensity and a slowdown in knowledge diffusion.*

Figure 4: Rising Software Expenditures and the Slowdown of Knowledge Diffusion



Note: The first panels show the logarithm of the software price from BEA. The second panel shows the firm-level software over knowledge capital ratio, averaged across all firms for all sectors in the Compustat dataset. The third panel shows firms' average upward citation as a fraction of total citations (in logs).

The third panel shows the rate of knowledge diffusion within 2-digit NAICS industries, measured as the intensity of upward citations from follower firms to industry leaders. We follow Akcigit and Ates (2021, 2023) to construct this measure by aggregating backward citations from firms with lower knowledge capital to their industry peers with higher knowledge capital. According to Akcigit and Ates (2023), a slowdown in citation flows from leaders to followers

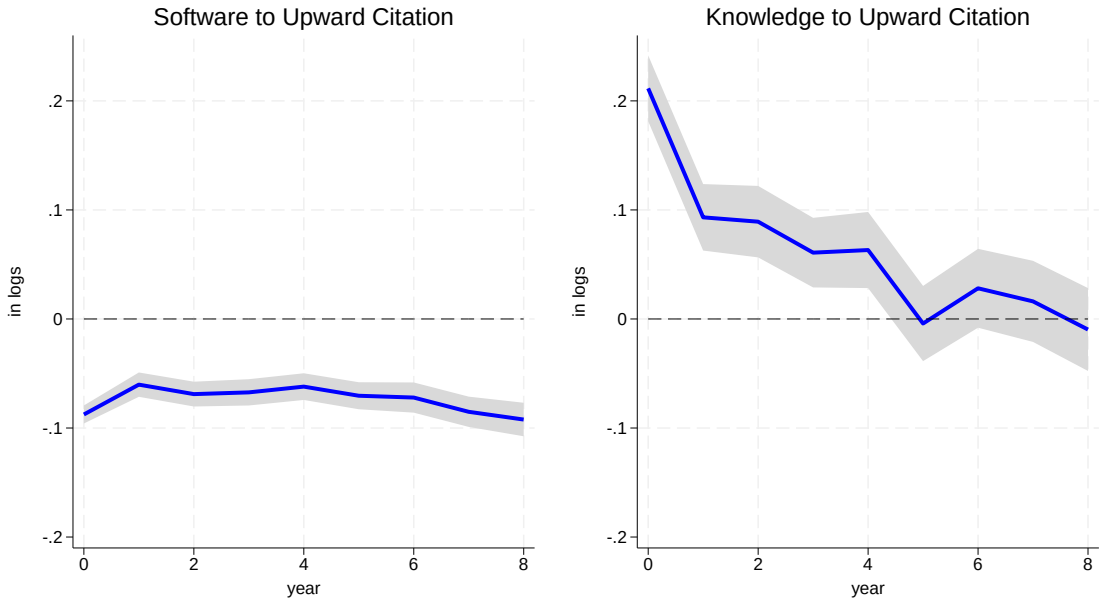
indicates reduced knowledge dissemination. The panel reveals that upward citations declined by approximately 12% since 2000, with most of the decline occurring after the Great Financial Crisis. We verify that this knowledge diffusion slowdown is robust across alternative diffusion measures. These patterns suggest that declining software costs have been accompanied by increased software investment and reduced knowledge diffusion.

Firm-level Evidence To highlight the role of software, we examine how changes in leading firms' software investment and knowledge capital affect citation flows from follower firms. The distinction between general R&D investment and software investment is particularly important, as software may create barriers to knowledge diffusion beyond its role in knowledge creation. We estimate the following firm-level local projection specification:

$$\Delta\text{citation}_{j,t+h} = \beta_{1,h}\Delta\text{software}_{j,t-1} + \beta_{2,h}\Delta\text{knowledge}_{j,t-1} + \beta_3 X_{j,t} + \alpha_j + \alpha_t + \epsilon_{j,t+h}, \quad (44)$$

where $\Delta\text{citation}_{j,t+h}$ denotes the change in the logarithm of upward citations received by firm j from period $t-1$ to $t+h$, with $h \in \{0, 1, \ldots, 15\}$. The variable $\Delta\text{software}_{i,t-1}$ measures the growth rate of software stock among leading firms—defined as firms within the same industry whose knowledge capital exceeds firm j's knowledge capital at time $t-1$—following the growth rate definition in Davis, Haltiwanger, and Schuh (1996). Similarly, $\Delta\text{knowledge}_{j,t-1}$ denotes the growth rate of knowledge capital among leading firms. We use growth rates rather than levels because the level variables are nonstationary. The specification includes firm fixed effects α_j and time fixed effects α_t. The vector $X_{j,t}$ contains three control variables: (i) the growth rate in the number of patents held by leading firms, which controls for the mechanical relationship that citation counts increase with the number of available patents; (ii) the total number of patents held by firm j, which controls for firm-specific citation patterns; and (iii) the change in average citations per patent for firm j, which accounts for potential changes in the sophistication of firm j's knowledge base.

Figure 5: Secrecy and Knowledge Diffusion at Firm-Level



Note: This figure shows the local projection of the changes of upward citation growth of firm j to changes in better firm's software and knowledge stock.

Figure 5 presents the estimated impulse response functions. The first panel shows the response of upward citations to changes in leading firms' software stock. A 1% increase in leading firms' software stock is associated with a 0.08% reduction in citation flows from leading firms to follower firms. This effect is statistically significant at the 1% level and persists over the entire horizon. The lower panel displays the response to changes in leading firms' knowledge capital, showing an opposite sign. This pattern suggests that software investment creates barriers to knowledge diffusion beyond the effects of knowledge creation itself, providing evidence for the differential impact of software investment relative to general R&D investment.[1] We summarize this finding as Fact 2.

Fact 2. *Software investment by leading firms slows the diffusion of knowledge to less productive firms.*

[1] A potential concern is whether the negative relationship between software investment and knowledge diffusion holds when examining a broader set of firms, including nonpublic firms not covered in Compustat. To address this concern, we use the EU-KLEMS database to construct sector-level measures of software inputs and the robustness check is presented in Appendix B.7.

3.2 The Extended Model

The empirical facts documented above establish a robust negative correlation between software investment and knowledge diffusion at the firm- and economy-level. Motivated by the two patterns, we extend the baseline model to allow firms to invest in both R&D and data&software, where the latter serves as an input for knowledge creation but simultaneously creates barriers to knowledge diffusion.

The extended model maintains the same general equilibrium structure as the baseline model developed in Section 2.2, with inventors innovating to sell ideas to production firms. The key extension is that inventors can now invest in two types of inputs: R&D investment $i \geq 0$ and data&software investment $d \geq 0$. Let $r_t > 0$ denote the cost of developing data&software at time t. The accumulation of data&software enhances knowledge creation productivity but creates barriers that impede knowledge diffusion. The firm's value function becomes:

$$W_t(q) = \max_{i,d} \left\{ -i - r_t d + M_t \frac{\kappa_t \left(q^\alpha i^\beta d^\gamma\right)^{\rho\theta} + \mu_t \overline{\mu}_t}{\kappa_t \overline{\kappa}_t + \mu_t \overline{\mu}_t} \int_0^\infty \left[\pi_{t+1}(\tilde{q}) + W_{t+1}(\tilde{q})\right] dF_{t+1}(\tilde{q}) \right\}, \quad (45)$$

where q denotes the firm's beginning-of-period quality, following the notation convention established in Section 2.2. As in the baseline model, firms create knowledge through either innovation or diffusion. Knowledge creation productivity now takes a Cobb-Douglas form in R&D investment i and data&software investment d: $q^\alpha i^\beta d^\gamma$, where $\beta, \gamma > 0$ and $\beta + \gamma = 1 - \alpha$ to maintain constant returns to scale in investment. This specification ensures that the production function for knowledge creation exhibits constant returns to scale in the investment inputs, while maintaining complementarity between existing quality and investment. The first-order conditions for optimal investment are:

$$i_t(q) = \left\{ \frac{\rho\theta\beta\left(\gamma\beta^{-1}\right)^{\rho\theta\gamma} \kappa_t \overline{W}_t r_t^{-\rho\theta\gamma}}{\left[1 - (1-\alpha)\rho\theta\right]\kappa_t \overline{\kappa}_t + \mu_t \overline{\mu}_t} \right\}^{\frac{1}{1-\rho\theta(1-\alpha)}} q^{\frac{\rho\theta\alpha}{1-\rho\theta(1-\alpha)}}, \quad (46)$$

$$d_t(q) = \frac{r_t^{-1}\gamma}{\beta} i_t(q), \quad (47)$$

where $\overline{W}_t \equiv \int_0^\infty W_{t+1}(\tilde{q}) dF_{t+1}(\tilde{q})$ denotes the average continuation value, following the notation

convention from the baseline model. The aggregate innovation intensity is given by

$$\overline{\kappa}_t = \int_0^\infty \left(q^\alpha i_t(q)^\beta d_t(q)^\gamma \right)^{\rho\theta} dF_t(q), \tag{48}$$

and the aggregate diffusion intensity is

$$\overline{\mu}_t = \int_0^\infty \left[1 - \xi_t(q, i_t(q), d_t(q))\right] \left(q^\alpha i_t(q)^\beta d_t(q)^\gamma \right)^{\rho\theta} dF_t(q), \tag{49}$$

where $\xi_t(q,i,d) \in [0,1]$ represents the barrier to knowledge diffusion that depends on the firm's quality and investment choices. When $\xi_t = 0$, there is no barrier and diffusion proceeds as in the baseline model; when $\xi_t = 1$, diffusion is completely blocked. The barrier function ξ_t is the key mechanism through which data&software investment affects knowledge diffusion.

The law of motion for the detrended Fréchet scale parameter $\widehat{\Lambda}_t$ becomes:

$$\begin{aligned}\widehat{\Lambda}_{t+1} = & \frac{\gamma^{\rho\theta\gamma}\beta^{\rho\theta\beta}\left(1+\tilde{\mu}_t\right) r_t^{-\rho\theta\gamma} g^{-\frac{1}{1-\alpha\rho-\frac{1-\nu}{\nu}\rho(1-\alpha)}}}{\left(\left[1-(1-\alpha)\rho\theta\right]+\tilde{\mu}_t\right)^{(1-\alpha)\rho\theta}} \times \\ & \widehat{\kappa}_t \widehat{W}_t^{\rho\theta(1-\alpha)} \widehat{\Lambda}_t^{\alpha\rho+\frac{1-\nu}{\nu}\rho(1-\alpha)} \left[\Gamma\left(1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right)\right]^{1-\rho\theta(1-\alpha)},\end{aligned} \tag{50}$$

where $\tilde{\mu}_t \equiv \mu_t \overline{\mu}_t / (\kappa_t \overline{\kappa}_t)$ measures the relative importance of diffusion versus innovation. This ratio is given by:

$$\tilde{\mu}_t = \frac{\widehat{\mu}_t}{\widehat{\kappa}_t} \int_0^1 \left[1 - \xi_t(q(Q))\right] \frac{(-\log Q)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}}}{\int_0^1 (-\log Q)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}} dQ} dQ, \tag{51}$$

where $q(Q) = F_t^{-1}(Q)$ denotes the quality level at quantile Q of the knowledge distribution. This expression shows that the relative importance of diffusion depends on the cross-sectional distribution of barriers across firms.

Barrier Functional Form The functional form of the firm-specific barrier $\xi_t(q,i,d)$ can be general, but to reconcile the empirical facts, it must be correlated with data&software investment d. By equations (46) and (47), optimal data&software investment is negatively related to r_t and positively related to q in equilibrium. For tractability and to ensure $\xi_t \in [0,1]$,

we posit the following exponential form:

$$\xi_t(q) = \exp\left(-\eta_t r_t^{\psi} q^{-\theta}\right), \tag{52}$$

where $\eta_t > 0$ is an aggregate component that shifts the barrier for all firms, and $\psi > 0$ governs the elasticity of the barrier with respect to the cost of data&software. This specification implies that barriers are higher for firms with greater quality (which invest more in data&software) and lower when the cost of data&software is higher. The exponential form ensures that $\xi_t \in [0, 1]$ for all parameter values, with $\xi_t \to 0$ as $q \to 0$ or $r_t \to \infty$, and $\xi_t \to 1$ as $q \to \infty$ or $r_t \to 0$. Substituting (52) into (51) yields:

$$\tilde{\mu}_t = \frac{\widehat{\mu}_t}{\widehat{\kappa}_t}\left[1 - \left(1 + \eta_t \Lambda_t^{-1} r_t^{\psi}\right)^{-\left(1 - \frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right)}\right]. \tag{53}$$

This expression shows that knowledge diffusion (relative to creation) slows down endogenously as the cost of data&software declines, consistent with the empirical patterns documented above. Specifically, a decline in r_t leads to increased data&software investment by firms, which raises barriers ξ_t and reduces the relative importance of diffusion $\tilde{\mu}_t$.

Economic Interpretation The extended model generates a tension between two opposing effects of declining data&software costs. First, a lower r_t directly incentivizes data&software investment, which enhances knowledge creation productivity and increases innovation. Second, by equation (53), the relative importance of diffusion $\tilde{\mu}_t$ is now endogenous to r_t: a lower r_t leads to higher barriers ξ_t (since firms invest more in data&software) and therefore lower diffusion $\tilde{\mu}_t$. This second effect can potentially dominate the first, leading to a negative relationship between data&software investment and aggregate knowledge growth despite increased innovation at the firm level. The slowdown in knowledge diffusion manifests as a lower rate of citing others and a higher rate of self-citation, consistent with the empirical patterns we document. This mechanism provides a microfoundation for the negative relationship between software investment and knowledge diffusion documented in the stylized facts. We now turn to quantitative analysis to illustrate these mechanisms.

3.3 Quantitative Analysis

This subsection presents the quantitative properties of the extended model through calibration and counterfactual analysis. We calibrate the model to match key moments from the firm-level data and then use the calibrated model to quantify how changes in the cost of data and software propagate through innovation and diffusion.

Model Parameterization We group the structural parameters into two categories: those assigned externally based on the literature and those calibrated to match data moments. The first group consists of δ, ρ, θ, ν, and ε. We set the discount factor $\delta = 0.90$. Following Cai et al. (2022), we set $\rho = 0.2$ and $\theta = 4.5$. We set $\nu = 1$ so that utility takes logarithmic form. Finally, we set $\varepsilon = 0.75$ following Broda and Weinstein (2006).

Table 1: Parameterization

Parameter	Symbol	Value
Discount factor	δ	0.90
Tail parameter (idea quality)	θ	4.50
Strength of idea generation	ρ	0.20
Utility curvature	ν	1.00
Elasticity parameter	ε	0.75
Knowledge share	α	0.75
R&D share	β	0.13
Data/software share	γ	0.12
Barrier elasticity w.r.t. r_t	ψ	1.48
Barrier shifter	η	0.76
Exogenous trend growth	g	1.032

Note: Assigned parameters follow the cited literature; remaining parameters are calibrated as described in the text. We normalize $\widehat{\kappa} = \widehat{\mu} = 1$.

The second group consists of α, β, γ, ψ, η, and g, which we calibrate to match data moments. We normalize the level of detrended innovation efficiency and meeting intensity by

setting $\widehat{\kappa} = \widehat{\mu} = 1$, so that the calibrated moments discipline the relative diffusion object $\tilde{\mu}_t$ and the barrier parameters.

The model implies a log-linear relationship between R&D investment and knowledge:

$$\log\left(\frac{i_t(q)}{\bar{i}_t}\right) = \frac{\rho\theta\alpha}{1-\rho\theta(1-\alpha)}\log\left(\frac{q}{\bar{q}_t}\right), \tag{54}$$

where $\bar{i}_t$ and $\bar{q}_t$ denote cross-sectional averages. We choose $\alpha = 0.75$ to match the empirical elasticity of approximately 0.87.[2] The model also implies that the ratio of data&software investment to R&D investment satisfies $r_t d_t(q)/i_t(q) = \gamma/\beta$ for a firm with beginning-of-period quality q. We use the average share of data&software investment relative to R&D investment in the data to pin down $\gamma = 0.12$, which implies $\beta = 0.13$ given the constant returns to scale restriction $\beta + \gamma = 1 - \alpha$.

To calibrate the parameters associated with diffusion barriers, we exploit the model's cross-sectional predictions for citation patterns. Let $o_t(Q)$ denote the probability that a firm at quantile Q of the knowledge distribution cites others' ideas rather than its own. The model implies:

$$o_t(Q) = \frac{\mu_t\overline{\mu}_t}{\kappa_t\left(q^\alpha i^\beta d^\gamma\right)^{\rho\theta} + \mu_t\overline{\mu}_t} = \frac{\tilde{\mu}_t}{\tilde{\mu}_t + (-\log Q)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}}\left[\Gamma\left(1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right)\right]^{-1}}. \tag{55}$$

Taking the log-odds ratio yields:

$$\log\left[\frac{o_t(Q)}{1-o_t(Q)}\right] = \log(\tilde{\mu}_t) + \frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\log(-\log Q) + \log\left[\Gamma\left(1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right)\right]. \tag{56}$$

Substituting the expression for $\tilde{\mu}_t$ from equation (53) and using the approximation $\log(1+x) \approx x$ for small x, we obtain:

$$\begin{aligned}\log\left[\frac{o_t(Q)}{1-o_t(Q)}\right] \approx{}& \psi\left[1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right]\log r_t + \frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\log(-\log Q) \\ &+ \log\left[\Gamma\left(1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right)\right] + \left[1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right]\log(\eta) + \log\left(\frac{\widehat{\mu}}{\widehat{\kappa}}\right),\end{aligned}$$

where we assume that $\widehat{\mu}$ and $\widehat{\kappa}$ are constant over time to identify ψ. Normalizing $\widehat{\mu} = \widehat{\kappa} = 1$

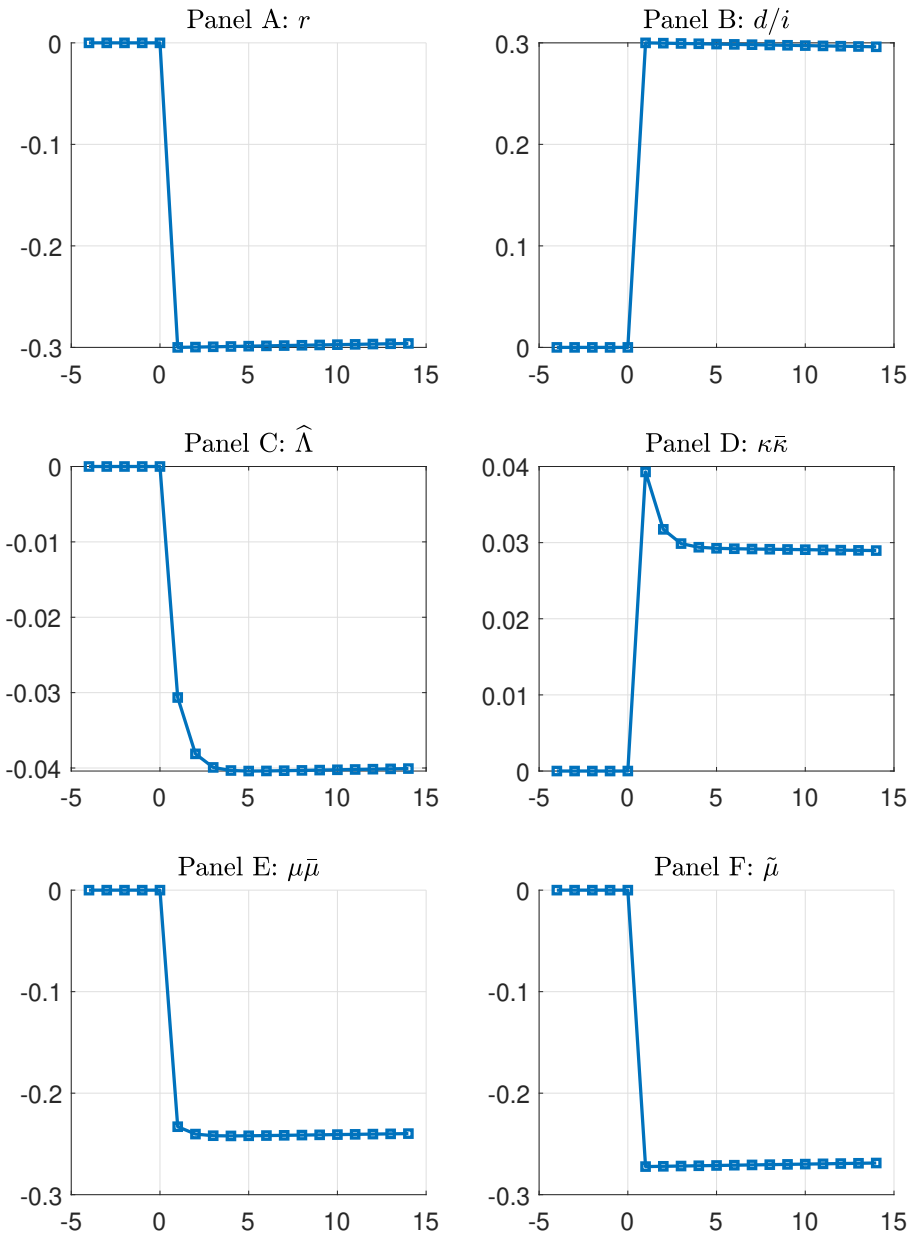
[2] Specifically, $\alpha = \varepsilon(1-\varepsilon)^{-1}\rho\theta(1-\rho\theta)^{-1}$, where ε is the elasticity estimated from the data.

without loss of generality, we use the fact that r_t declined by 25% and the log-odds of citing others declined by approximately 30% in our data to pin down $\psi = 1.48$. We then use the level of diffusion to calibrate $\eta = 0.76$. This calibration strategy exploits both the time-series variation in software costs and the cross-sectional variation in citation patterns across firms of different quality levels.

For the growth rate g, the model implies $\log g = \theta \left[1 - \alpha\rho - \frac{1-\nu}{\nu}(1-\alpha)\rho\right] \log(g_{\bar{q}})$, where $g_{\bar{q}}$ is the average growth rate of total factor productivity. Using $g_{\bar{q}} = 1.01$ from U.S. data, we obtain $g = 1.032$. This growth rate is consistent with the long-run growth rate of aggregate productivity in the U.S. economy.

Reconciling Fact 1 Figure 6 presents the impulse response functions to a one-time 30% permanent decrease in the cost of data&software (r_t). Panel B shows the growth rate of knowledge capital, Panel C displays knowledge creation, Panel D shows knowledge diffusion, Panel E presents diffusion component, and Panel F shows the relative diffusion measure $\tilde{\mu}_t$. The decline in r_t generates an immediate but short-lived innovation boom, as firms respond to lower costs by increasing both R&D and data&software investment. However, in the long run, innovation declines as diffusion slows down and the knowledge distribution endogenously shifts. This pattern illustrates the tension between the positive effect of lower costs on knowledge creation and the negative effect on knowledge diffusion. The quantitative results demonstrate that the negative diffusion effect can dominate the positive creation effect, leading to slower aggregate growth despite increased firm-level innovation.

Figure 6: Impact of Data&Software Cost Reduction



Note: This figure shows the impact of a one-time 30% permanent decrease in the cost of data&software.

Note that the key qualitative prediction—that lower r_t can raise short-run innovation yet reduce long-run growth by weakening diffusion—is robust to reasonable variation in the barrier elasticity ψ and shifter η. Increasing ψ strengthens the endogenous decline in diffusion in response to lower costs, while decreasing η weakens it. We report additional sensitivity exercises

in the Appendix A.8.

Reconciling Fact 2 We now demonstrate that the model is qualitatively consistent with the differential response of upward citations to leading firms' R&D and data&software investment, as documented in the stylized facts. Consider a firm at quantile Q of the knowledge distribution, and suppose we engineer shocks to the investment choices of all firms with quantiles $Q' > Q$. Specifically, we shift each leading firm's R&D investment by $100 \times \Delta_i$ percent and data&software investment by $100 \times \Delta_d$ percent. The knowledge available from leading firms becomes:

$$\overline{\mu}_{ut}\left(Q;\Delta_i,\Delta_d\right) = \int_Q^1 \left[1 - e^{-\widehat{\eta}_t(1-\Delta_d)^{\chi} r_t^{\psi} q^{-\theta}}\right] \left[q^{\alpha}\left(1+\Delta_i\right)^{\beta} i_t\left(q\right)^{\beta}\left(1+\Delta_d\right)^{\gamma} d_t\left(q\right)^{\gamma}\right]^{\rho\theta} dF_t\left(q\right), \tag{57}$$

where $\chi > 0$ governs the sensitivity of barriers to additional data&software investment. The knowledge available from follower firms (with quantiles $Q' < Q$) remains unchanged:

$$\overline{\mu}_{dt}\left(Q\right) = \int_0^Q \left[1 - e^{-\widehat{\eta}_t r_t^{\psi} q^{-\theta}}\right] \left(q^{\alpha} i_t\left(q\right)^{\beta} d_t\left(q\right)^{\gamma}\right)^{\rho\theta} dF_t\left(q\right). \tag{58}$$

The fraction of citations to leading firms is then:

$$u_t\left(Q;\Delta_i,\Delta_d\right) = \frac{\mu_t \overline{\mu}_{ut}\left(Q;\Delta_i,\Delta_d\right)}{\kappa_t \left(q^{\alpha} i_t\left(q\right)^{\beta} d_t\left(q\right)^{\gamma}\right)^{\rho\theta} + \mu_t\left[\overline{\mu}_{ut}\left(Q;\Delta_i,\Delta_d\right) + \overline{\mu}_{dt}\left(Q\right)\right]}. \tag{59}$$

Proposition 4. *For all $0 < Q < 1$, we have*

$$\frac{\partial}{\partial \Delta_i} u_t\left(Q;\Delta_i,\Delta_d\right) > 0.$$

In addition, when χ is sufficiently large, we have

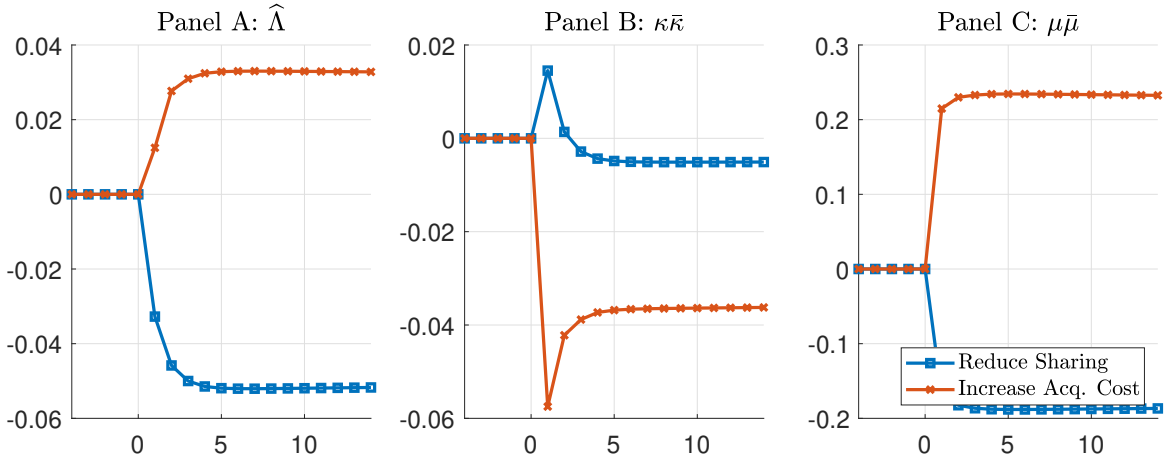
$$\frac{\partial}{\partial \Delta_d} u_t\left(Q;\Delta_i,\Delta_d\right) < 0.$$

The proof follows from differentiating equation (59) with respect to Δ_i and Δ_d, and noting that an increase in R&D investment increases the quality of knowledge available from leading firms, while an increase in data&software investment increases barriers when χ is sufficiently large.

Economic Interpretation This proposition formalizes the differential response of upward citations to R&D and data&software investment documented in the stylized facts. An increase in leading firms' R&D investment ($\Delta_i > 0$) increases upward citations, as it enhances the quality of knowledge available from above without creating additional barriers to diffusion. In contrast, an increase in leading firms' data&software investment ($\Delta_d > 0$) decreases upward citations when ξ is sufficiently large, because additional data&software investment creates stronger barriers to knowledge diffusion that offset the positive effect of enhanced knowledge quality. This pattern aligns with the empirical findings in Figure 5, where software investment by leading firms reduces citation flows to follower firms, while R&D investment has a negligible or mildly negative effect.

Implications for Privacy and Data Sharing Our model distinguishes between two dimensions of data sharing that have different implications for aggregate growth. The variable r_t captures the cost of data acquisition from consumers, which relates to consumer privacy protection: higher privacy protection corresponds to higher r_t. The parameter $\widehat{\eta}_t$ captures firms' willingness to share data with other firms, which relates to industrial data sharing: lower $\widehat{\eta}_t$ corresponds to reduced data sharing between firms. Figure 7 presents impulse response functions to permanent shocks to these two variables. The results show that while consumer privacy protection (higher r_t) can increase growth by reducing barriers to diffusion, industrial data sharing (higher $\widehat{\eta}_t$) is overall beneficial for aggregate growth, as it facilitates knowledge diffusion across firms. This distinction highlights the different policy implications of consumer privacy regulations versus industrial data sharing policies.

Figure 7: Privacy and Data Sharing



Note: Reduced data sharing corresponds to a negative shock to $\widehat{\eta}_t$ in the barrier function. An increase in the cost of acquisition corresponds to a positive shock to r_t. Both shocks are 30% deviations from their average and permanent.

4 Conclusion

This paper studies how firm-specific intangible capital shapes the trade-off between knowledge creation and knowledge diffusion. We develop a dynamic general equilibrium model in which firms innovate and learn from others, but investments in proprietary data and software both raise the productivity of idea creation and reduce the transmissibility of the resulting knowledge. The model remains tractable because an extreme-value structure preserves a Fréchet distribution of idea quality and yields a closed-form law of motion for the frontier.

In the data, we document that software accumulation by leading firms is associated with a persistent decline in vertical knowledge diffusion: a 1% increase in leaders' software stock reduces upward citations from followers by 0.08%, while changes in leaders' R&D have little effect. Calibrated to these moments, the model reproduces the post-2000 decline in diffusion and implies that falling software costs can raise innovation incentives in the short run yet lower long-run growth by reducing diffusion.

The framework highlights a policy tension between promoting firm-specific intangible investment and preserving knowledge spillovers. Policies that increase the cost of acquiring pro-

prietary data (e.g., stronger consumer privacy protections) can increase growth in the model by limiting diffusion barriers, while policies that increase inter-firm data sharing can raise growth by directly strengthening diffusion. More broadly, the mechanism is likely to matter in settings where leading firms accumulate nontransferable inputs—including in AI, where proprietary training data and model parameters can amplify private returns to creation while limiting diffusion.

References

Akcigit, U. and Ates, S. T. (2021). Ten facts on declining business dynamism and lessons from endogenous growth theory. *American Economic Journal: Macroeconomics*, 13(1):257–98.

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

Akcigit, U., Celik, M. A., and Greenwood, J. (2016). Buy, keep, or sell: Economic growth and the market for ideas. *Econometrica*, 84(3):943–984. Publisher: Wiley Online Library.

Broda, C. and Weinstein, D. E. (2006). Globalization and the gains from variety. *Quarterly Journal of Economics*, 121(2):541–585.

Buera, F. J. and Oberfield, E. (2020). The global diffusion of ideas. *Econometrica*, 88(1):83–114.

Cai, S., Caliendo, L., Parro, F., and Xiang, W. (2022). Mechanics of spatial growth. Technical report, National Bureau of Economic Research.

Corrado, C., Haskel, J., Jona-Lasinio, C., and Iommi, M. (2022). Intangible capital and modern economies. *Journal of Economic Perspectives*, 36(3):3–28.

Crouzet, N., Eberly, J. C., Eisfeldt, A. L., and Papanikolaou, D. (2022). The economics of intangible capital. *Journal of Economic Perspectives*, 36(3):29–52.

Davis, S. J., Haltiwanger, J., and Schuh, S. (1996). Small business and job creation: Dissecting the myth and reassessing the facts. *Small Business Economics*, 8(4):297–315.

Dorn, D., Hanson, G. H., Pisano, G., and Shu, P. (2020). Foreign competition and domestic innovation: Evidence from US patents. *American Economic Review: Insights*, 2(3):357–74.

Garner, C., Russell, M., Bessen, J., Meyer, P. B., and Sveikauskas, L. (2021). Intangible capital and us productivity growth in 61 industries.

Haskel, J. and Westlake, S. (2018). *Capitalism without Capital: The Rise of the Intangible Economy*. Princeton University Press, Princeton, New Jersey.

Hopenhayn, H., Shi, L., and Okumura, K. (2025). Knowledge creation and diffusion with limited appropriation. Working paper.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–58.

König, M. D., Lorenz, J., and Zilibotti, F. (2016). Innovation vs. imitation and the evolution of productivity distributions. *Theoretical Economics*, 11(3):1053–1102.

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

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

Peters, R. H. and Taylor, L. A. (2007). Intangible capital and the investment-q relation. page 52.

PwC (2021). Software costs. `https://viewpoint.pwc.com/dt/us/en/pwc/accounting_guides/software/software/about_this_guide.html` Accessed: Date of access.

Veldkamp, L. and Chung, C. (2019). Data and the aggregate economy. *Journal of Economic Literature*.

Appendices

A Model Appendix **42**

A.1 The Demand and Profit Functions 42

A.2 The Investment Decision 43

A.3 Proof of Proposition 1 43

A.4 Proof of Corollary 2 44

A.5 A Knife-edge Case 46

A.6 Data and Software Application 47

A.7 Proof of Proposition 4 48

A.8 Sensitivity Analysis 49

A.9 Construction of Model and Data Moments 49

A.10 Vertical Knowledge Diffusion 49

B Data Appendix **50**

B.1 Matching Assignees of PatentsView to Compustat 50

B.2 Note on the Measurement of Software 51

B.3 Alternative Measures of Diffusion 52

B.4 Internal Software as a Measure of Firm Secrecy 53

B.5 Citation Flows as a Measure of Knowledge Diffusion 53

B.6 Summary Statistics 54

B.7 Sector-level Results 57

A Model Appendix

A.1 The Demand and Profit Functions

This subsection collects the demand and profit expressions used in the main text. The steps are standard; we omit intermediate algebra and report the key relationships.

The household maximizes utility taking prices as given:

$$\max_{c_t(\omega),H_t} u\left\{\left[\int \left(q_t(\omega)c_t(\omega)\right)^{\varepsilon} d\omega\right]^{\frac{1}{\varepsilon}}\right\} - H_t$$
$$s.t \int p_t(\omega)c_t(\omega)d\omega \leq w_t H_t$$

The first-order conditions imply $\lambda_t w_t = 1$ and the standard demand schedule:

$$c_t(\omega) = \left[\varepsilon \frac{u'(C_t)}{v'(H_t)}\right]^{\frac{1}{1-\varepsilon}} q_t(\omega)^{\frac{\varepsilon}{1-\varepsilon}} C_t.$$

The producer's profit function is

$$\frac{\pi_t(\omega)}{w_t} = \max_{h_t(\omega)} \frac{u'(C_t)}{v'(H_t)} C_t^{1-\varepsilon} q_t(\omega)^{\varepsilon} h_t(\omega)^{\varepsilon} - h_t(\omega),$$

which implies profits are proportional to quality:

$$\pi_t(\omega) = (1-\varepsilon)\,\varepsilon^{\frac{\varepsilon}{1-\varepsilon}} \left(\frac{u'(C_t)}{v'(H_t)}\right)^{\frac{1}{1-\varepsilon}} C_t\, q_t(\omega)^{\frac{\varepsilon}{1-\varepsilon}}.$$

With $u(C_t) = C_t^{1-\nu}/(1-\nu)$ and $v(H_t) = H$, and we choose $w_t = 1$,

$$p_t(\omega) = \frac{1}{\lambda_t} C_t^{1-\varepsilon-\nu} q_t(\omega)^{\varepsilon} c_t(\omega)^{-(1-\varepsilon)},$$
$$\pi_t(\omega) = (1-\varepsilon)\,\varepsilon^{\frac{1-\nu}{\nu}} \left(\int q_t(\omega)^{\frac{\varepsilon}{1-\varepsilon}} d\omega\right)^{\frac{1-\varepsilon}{\varepsilon}\left(\frac{1}{\nu}-\frac{1}{1-\varepsilon}\right)} q_t(\omega)^{\frac{\varepsilon}{1-\varepsilon}}.$$

Finally, the stochastic discount factor M_t used in the inventor's problem satisfies

$$M_t \equiv \delta \frac{u'(C_{t+1})}{u'(C_t)} = \delta \left(\frac{\Lambda_{t+1}}{\Lambda_t}\right)^{-\frac{1}{\theta}},$$

where the last equality uses the Fréchet aggregation implied by $F_t(q) = \exp(-\Lambda_t q^{-\theta})$.

A.2 The Investment Decision

The first order condition for i is given by:

$$i = M_t \frac{(1-\alpha)\,\rho\theta\kappa_t \left(q_t^{\alpha} i_t^{1-\alpha}\right)^{\rho\theta}}{\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t} \int_0^{\infty} \left[\pi_{t+1}(\tilde{q}) + W_{t+1}(\tilde{q})\right] dF_{t+1}(\tilde{q}) \tag{60}$$

Aggregating the value function across firms yields a recursion for $\overline{W}_t \equiv \int_0^{\infty} W_t(q)\, dF_t(q)$:

$$\overline{W}_t = M_t \left[1 - (1-\alpha)\,\rho\theta \frac{\kappa_t\overline{\kappa}_t}{\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t}\right] \left\{(1-\varepsilon)\,\varepsilon^{\frac{1-\nu}{\nu}} \Lambda_{t+1}^{\frac{1}{\theta}\frac{1-\nu}{\nu}} \Gamma\left(1 - \frac{1}{\theta}\frac{\varepsilon}{1-\varepsilon}\right)^{\frac{1-\varepsilon}{\varepsilon}\frac{1-\nu}{\nu}} + \overline{W}_{t+1}\right\} \tag{61}$$

Combining (60) and (61) yields the closed-form policy function reported in the main text.

A.3 Proof of Proposition 1

The aggregate knowledge creation term is $\overline{\kappa}_t \equiv \int_0^{\infty} \left(q^{\alpha} i_t(q)^{1-\alpha}\right)^{\rho\theta} dF_t(q)$. Substituting the optimal policy $i_t(q)$ and simplifying yields: define $\tilde{\mu}_t \equiv \mu_t\overline{\mu}_t / \kappa_t\overline{\kappa}_t$, the $\overline{\kappa}_t$ term can be represented as

$$\overline{\kappa}_t = \left\{\frac{(1-\alpha)\,\rho\theta\overline{W}_t}{[1-(1-\alpha)\,\rho\theta] + \tilde{\mu}_t}\right\}^{(1-\alpha)\rho\theta} \Lambda_t^{\alpha\rho} \left[\int_0^1 (-\log Q)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}}\, dQ\right]^{1-\rho\theta(1-\alpha)} \tag{62}$$

where we use $Q \equiv e^{-\Lambda_t q^{-\theta}}$ so that $\Lambda_t^{\frac{1}{\theta}} (-\log Q)^{-\frac{1}{\theta}} = q$. Next, by the definition of $\overline{\mu}_t$:

$$\overline{\mu}_t = \int_0^{\infty} \left[1 - \xi_t(q)\right] q^{\alpha\rho\theta} \left[\left\{\frac{(1-\alpha)\,\rho\theta\kappa_t\overline{W}_t}{[1-(1-\alpha)\,\rho\theta]\,\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t}\right\}^{\frac{1}{1-(1-\alpha)\rho\theta}} q_t^{\frac{\alpha\rho\theta}{1-(1-\alpha)\rho\theta}}\right]^{(1-\alpha)\rho\theta} dF_t(q),$$

and note the relationship between $\overline{\mu}_t$ and $\overline{\kappa}_t$:

$$\overline{\mu}_t = \overline{\kappa}_t \int_0^1 \left[1 - \xi_t\left(F_t^{-1}(Q)\right)\right] \frac{(-\log Q)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}}}{\int_0^1 (-\log Q)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}}\, dQ} dQ.$$

Divide both sides of

$$\overline{W}_t = M_t \left[1 - (1-\alpha)\rho\theta \frac{\kappa_t \overline{\kappa}_t}{\kappa_t \overline{\kappa}_t + \mu_t \overline{\mu}_t}\right] \left[(1-\varepsilon)\varepsilon^{\frac{1-\nu}{\nu}} \Lambda_{t+1}^{\frac{1}{\theta}\frac{1-\nu}{\nu}} \Gamma\left(1 - \frac{\varepsilon}{\theta(1-\varepsilon)}\right)^{\frac{1-\varepsilon}{\varepsilon}\frac{1-\nu}{\nu}} + \overline{W}_{t+1}\right]$$

by $\Lambda_t^{\frac{1}{\theta}\frac{1-\nu}{\nu}}$ and define $\widehat{W}_t = \overline{W}_t / \Lambda_t^{\frac{1}{\theta}\frac{1-\nu}{\nu}}$:

$$\widehat{W}_t = \delta \left(\frac{\Lambda_{t+1}^*}{\Lambda_t^*} \frac{\widehat{\Lambda}_{t+1}}{\widehat{\Lambda}_t}\right)^{\frac{1}{\theta}\frac{1-2\nu}{\nu}} \left(1 - \frac{\rho\theta(1-\alpha)}{1+\tilde{\mu}_t}\right) \left[(1-\varepsilon)\varepsilon^{\frac{1-\nu}{\nu}} \Gamma\left(1 - \frac{\varepsilon}{\theta(1-\varepsilon)}\right)^{\frac{1-\varepsilon}{\varepsilon}\frac{1-\nu}{\nu}} + \widehat{W}_{t+1}\right],$$

and in the case of balanced growth

$$\widehat{W}_t = \delta \left(g_\Lambda \frac{\widehat{\Lambda}_{t+1}}{\widehat{\Lambda}_t}\right)^{\frac{1}{\theta}\frac{1-2\nu}{\nu}} \left[1 - \frac{(1-\alpha)\rho\theta}{1+\tilde{\mu}_t}\right] \left[(1-\varepsilon)\varepsilon^{\frac{1-\nu}{\nu}} \Gamma\left(1 - \frac{1}{\theta}\frac{\varepsilon}{1-\varepsilon}\right)^{\frac{1-\varepsilon}{\varepsilon}\frac{1-\nu}{\nu}} + \widehat{W}_{t+1}\right]. \quad (63)$$

clearly $\widehat{W}_t$ eventually converges to

$$\lim_{t\to\infty} \widehat{W}_t = \frac{\delta g_\Lambda^{\frac{1}{\theta}\frac{1-2\nu}{\nu}} \left[1 - \frac{(1-\alpha)\rho\theta}{1+\tilde{\mu}_\infty}\right] \left[(1-\varepsilon)\varepsilon^{\frac{1-\nu}{\nu}} \Gamma\left(1 - \frac{1}{\theta}\frac{\varepsilon}{1-\varepsilon}\right)^{\frac{1-\varepsilon}{\varepsilon}\frac{1-\nu}{\nu}}\right]}{1 - \delta g_\Lambda^{\frac{1}{\theta}\frac{1-2\nu}{\nu}} \left[1 - \frac{(1-\alpha)\rho\theta}{1+\tilde{\mu}_\infty}\right]}.$$

The law of motion for $\Lambda_{t+1} = \kappa_t \overline{\kappa}_t + \mu_t \overline{\mu}_t$ and the detrended system follow by substituting the expressions for $\overline{\kappa}_t$ and $\overline{\mu}_t$ into the decomposition in Proposition 3. The transversality condition is $\delta < g_\Lambda^{\frac{1}{\theta}\frac{2\nu-1}{\nu}}$.

A.4 Proof of Corollary 2

Let $\widehat{w}_t$, $\widehat{\lambda}_t$ denote the log-deivations of $\widehat{W}_t$ and $\widehat{\Lambda}_t$. Assume $\widehat{w}_{t+k} = 0$, $\forall k \geq 1$. Linearize the dynamics of $\widehat{W}_t$ gives

$$\widehat{w}_t = \frac{1}{\theta}\frac{1-2\nu}{\nu}\widehat{\lambda}_{t+1} + \frac{\rho\theta(1-\alpha)}{1-\rho\theta(1-\alpha)+\tilde{\mu}_t}\widehat{1+\tilde{\mu}_t}$$

and log-linearize the law of motion for $\widehat{\Lambda}_{t+1}$ gives

$$\widehat{\lambda}_{t+1} = \left[1 - \frac{(1-\alpha)\rho\theta(1+\widetilde{\mu}_t)}{1-(1-\alpha)\rho\theta+\widetilde{\mu}_t}\right]\widehat{1+\widetilde{\mu}_t} + \widehat{\widehat{\kappa}}_t + \rho\theta(1-\alpha)\widehat{w}_t$$

combine the two equations gives how diffusion shift the value function:

$$\begin{aligned}&\left[1-\frac{1-2\nu}{\nu}\rho\left(1-\alpha\right)\right]\widehat{w}_t\\&=\left\{\frac{1}{\theta}\frac{1-2\nu}{\nu}\left[1-\frac{\left(1-\alpha\right)\rho\theta\left(1+\widetilde{\mu}\right)}{1-\left(1-\alpha\right)\rho\theta+\widetilde{\mu}}\right]+\frac{\rho\theta\left(1-\alpha\right)}{1-\rho\theta\left(1-\alpha\right)+\tilde{\mu}}\right\}\widehat{1+\widetilde{\mu}_t}+\frac{1}{\theta}\frac{1-2\nu}{\nu}\widehat{\widehat{\kappa}}_t\end{aligned}$$

where $\widehat{1+\widetilde{\mu}_t}$ is the log-deviation of $1+\widetilde{\mu}_t$. By

$$\overline{\kappa}_t=\left(\frac{\rho\theta\left(1-\alpha\right)\kappa_t\overline{W}_t}{\left[1-\left(1-\alpha\right)\rho\theta\right]\kappa_t\overline{\kappa}_t+\mu_t\overline{\mu}_t}\right)^{\frac{\rho\theta(1-\alpha)}{1-(1-\alpha)\rho\theta}}\int_0^\infty\widetilde{q}^{\frac{\alpha\rho\theta}{1-(1-\alpha)\rho\theta}}dF_t\left(\widetilde{q}\right),$$

we have

$$\begin{aligned}\widehat{\kappa_t\overline{\kappa}_t}&=\left[1-\left(1-\alpha\right)\rho\theta\right]\left\{\frac{1}{1-\left(1-\alpha\right)\rho\theta}\widehat{\widehat{\kappa}}_t+\frac{\rho\theta\left(1-\alpha\right)}{1-\left(1-\alpha\right)\rho\theta}\left[\widehat{w}_t-\frac{1+\tilde{\mu}}{\left[1-\left(1-\alpha\right)\rho\theta\right]+\tilde{\mu}}\widehat{1+\widetilde{\mu}_t}\right]\right\}\\&=\widehat{\widehat{\kappa}}_t+\rho\theta\left(1-\alpha\right)\left[\widehat{w}_t-\frac{1+\tilde{\mu}}{\left[1-\left(1-\alpha\right)\rho\theta\right]+\tilde{\mu}}\widehat{1+\widetilde{\mu}_t}\right]\end{aligned}$$

Plugging in $\widehat{w}_t$, we have diffusion (i.e., $1+\widetilde{\mu}_t$) negatively impact on knowledge creation if and only if

$$\frac{1}{\theta}\frac{1-2\nu}{\nu}\left[1-\frac{\left(1-\alpha\right)\rho\theta\left(1+\widetilde{\mu}\right)}{1-\left(1-\alpha\right)\rho\theta+\widetilde{\mu}}\right]+\frac{\rho\theta\left(1-\alpha\right)}{1-\rho\theta\left(1-\alpha\right)+\tilde{\mu}}\leq\frac{\left(1-\frac{1-2\nu}{\nu}\rho\left(1-\alpha\right)\right)\left(1+\tilde{\mu}\right)}{\left[1-\left(1-\alpha\right)\rho\theta\right]+\tilde{\mu}}$$

$\widetilde{\mu}_t$ will reduce $\widehat{\kappa_t\overline{\kappa}_t}$. For example, in the case of $\nu=0.5$, the above inequality is $\widetilde{\mu}\geq 1-\rho\theta\left(1-\alpha\right)$, a condition that always hold.

In addition, note the law of motion of $\widehat{\Lambda}$ at the stady state is

$$\widehat{\Lambda}^{1-\alpha\rho-\frac{1-\nu}{\nu}\rho(1-\alpha)}\propto\left(1+\widetilde{\mu}\right)\left(\left[1-\left(1-\alpha\right)\rho\theta\right]+\widetilde{\mu}\right)^{-(1-\alpha)\rho\theta}\widehat{\kappa}\widehat{W}^{\rho\theta(1-\alpha)}\tag{64}$$

$$\equiv f\left(\tilde{\mu}\right)\tag{65}$$

where we ignore the constants and define the key objects as $f\left(\widetilde{\mu}\right)$. Then by noting the rela-

tionship between $\widehat{W}$ and $\widetilde{\mu}$ from equation (63):

$$\frac{d\widehat{W}}{d\tilde{\mu}} = \delta g_{\Lambda}^{\frac{1}{\theta}\frac{1-2\nu}{\nu}} \frac{(1-\varepsilon)\,\varepsilon^{\frac{1-\nu}{\nu}}\Gamma\left(1-\frac{1}{\theta}\frac{\varepsilon}{1-\varepsilon}\right)^{\frac{1-\varepsilon}{\varepsilon}\frac{1-\nu}{\nu}}}{\left(1-\delta g_{\Lambda}^{\frac{1}{\theta}\frac{1-2\nu}{\nu}}\left[1-\frac{(1-\alpha)\rho\theta}{1+\widetilde{\mu}}\right]\right)^2} \frac{(1-\alpha)\,\rho\theta}{(1+\tilde{\mu})^2}$$

we have

$$\begin{aligned}
\frac{f'(\tilde{\mu})}{f(\tilde{\mu})} &= \frac{1}{1+\widetilde{\mu}} - \frac{(1-\alpha)\,\rho\theta}{[1-(1-\alpha)\,\rho\theta]+\widetilde{\mu}} + \frac{1+\widetilde{\mu}}{1-(1-\alpha)\,\rho\theta+\widetilde{\mu}} \frac{1}{1-\delta g_{\Lambda}^{\frac{1}{\theta}\frac{1-2\nu}{\nu}}\left[1-\frac{(1-\alpha)\rho\theta}{1+\widetilde{\mu}}\right]} \frac{[(1-\alpha)\,\rho\theta]^2}{(1+\tilde{\mu})^2} \\
&\geq \frac{1}{1+\widetilde{\mu}_t} - \frac{(1-\alpha)\,\rho\theta}{[1-(1-\alpha)\,\rho\theta]+\widetilde{\mu}_t} + \frac{1+\widetilde{\mu}_t}{1-(1-\alpha)\,\rho\theta+\widetilde{\mu}_t} \frac{[(1-\alpha)\,\rho\theta]^2}{(1+\tilde{\mu}_t)^2} \\
&= \frac{1-(1-\alpha)\,\rho\theta}{1+\widetilde{\mu}_t} \\
&\geq 0
\end{aligned}$$

so that increase diffusion always lead increase in $\widehat{\Lambda}$ through $f(\widetilde{\mu})$.

A.5 A Knife-edge Case

By equation (63), we can characterize the dynamics in the knife-edge case $\alpha\rho+\frac{1-\nu}{\nu}\rho(1-\alpha)=0$:

$$\widehat{W}_t = \delta\left(g_{\Lambda,t}\right)^{\frac{1}{\theta}\frac{1-2\nu}{\nu}}\left(1-\frac{\rho\theta(1-\alpha)}{1+\tilde{\mu}_t}\right)\left[(1-\varepsilon)\,\varepsilon^{\frac{1-\nu}{\nu}}\Gamma\left(1-\frac{\varepsilon}{\theta(1-\varepsilon)}\right)^{\frac{1-\varepsilon}{\varepsilon}\frac{1-\nu}{\nu}}+\widehat{W}_{t+1}\right] \tag{66}$$

and also note the dynamics of knowledge:

$$g_{\Lambda t} = \frac{[\rho\theta(1-\alpha)]^{(1-\alpha)\rho\theta}(1+\tilde{\mu}_t)}{([1-(1-\alpha)\,\rho\theta]+\tilde{\mu}_t)^{(1-\alpha)\rho\theta}}\widehat{\kappa}_t\widehat{W}_t^{\rho\theta(1-\alpha)}\left[\Gamma\left(1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right)\right]^{1-\rho\theta(1-\alpha)} \tag{67}$$

gives two equation to solve for $\left\{g_{\Lambda t},\widehat{W}_t\right\}$. There is no closed form solution for g_Λ in general, but it is clear that the level of diffusion $\tilde{\mu}$ affects the growth rate g_Λ.

A.6 Data and Software Application

The first order condition for i and d is given by:

$$i = \rho\theta\beta M_t \frac{\kappa_t \left(q^\alpha i^\beta d^\gamma\right)^{\rho\theta}}{\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t} \int_0^\infty \left[\pi_{t+1}\left(\tilde{q}\right) + W_{t+1}\left(\tilde{q}\right)\right] dF_{t+1}\left(\tilde{q}\right),$$
$$r_t d = \rho\theta\gamma M_t \frac{\kappa_t \left(q^\alpha i^\beta d^\gamma\right)^{\rho\theta}}{\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t} \int_0^\infty \left[\pi_{t+1}\left(\tilde{q}\right) + W_{t+1}\left(\tilde{q}\right)\right] dF_{t+1}\left(\tilde{q}\right),$$

Solving these first-order conditions yields:

$$i = \left\{\frac{\rho\theta\beta^{1-\rho\theta\gamma}\gamma^{\rho\theta\gamma}\kappa_t\overline{W}_t}{\left[1-(1-\alpha)\rho\theta\right]\kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t}\right\}^{\frac{1}{1-\rho\theta(1-\alpha)}} r_t^{-\frac{\rho\theta\gamma}{1-\rho\theta(1-\alpha)}} q^{\frac{\rho\theta\alpha}{1-\rho\theta(1-\alpha)}}$$
$$d = \frac{r_t^{-1}\gamma}{\beta} i$$

and the aggregate knowledge creation term simplifies to

$$\overline{\kappa}_t = \gamma^{\rho\theta\gamma}\beta^{\rho\theta\beta}\left\{\frac{\rho\theta\overline{W}_t}{\left[1-(1-\alpha)\rho\theta\right]+\tilde{\mu}_t}\right\}^{\rho\theta(1-\alpha)} r_t^{-\rho\theta\gamma}\left[\int_0^\infty q^{\frac{\alpha\rho\theta}{1-\rho\theta(1-\alpha)}} dF_t\left(q\right)\right]^{1-\rho\theta(1-\alpha)}.$$

Similarly, the diffusion term is

$$\overline{\mu}_t = \overline{\kappa}_t \int_0^1 \left[1-\xi_t\left(F_t^{-1}\left(Q\right)\right)\right] \frac{\left(-\log Q\right)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}}}{\int_0^1 \left(-\log Q\right)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}} dQ} dQ.$$

where $\tilde{\mu}_t \equiv \mu_t\overline{\mu}_t/\kappa_t\overline{\kappa}_t$. The dynamics of value function can be similarly derived and the dynamics of dynamics of knowledge is given by:

$$\begin{aligned}
\Lambda_{t+1} &= \kappa_t\overline{\kappa}_t + \mu_t\overline{\mu}_t \\
&= \frac{\gamma^{\rho\theta\gamma}\beta^{\rho\theta\beta}\left(1+\tilde{\mu}_t\right) r_t^{-\rho\theta\gamma} g^{-\frac{1}{1-\alpha\rho-\frac{1-\nu}{\nu}\rho(1-\alpha)}}}{\left(\left[1-(1-\alpha)\rho\theta\right]+\tilde{\mu}_t\right)^{(1-\alpha)\rho\theta}} \widehat{\kappa}_t \widehat{W}_t^{\rho\theta(1-\alpha)} \widehat{\Lambda}_t^{\alpha\rho+\frac{1-\nu}{\nu}\rho(1-\alpha)} \times \\
&\left[\Gamma\left(1-\frac{\alpha\rho}{1-\rho\theta\left(1-\alpha\right)}\right)\right]^{1-\rho\theta(1-\alpha)}
\end{aligned}$$

where using the barrier functional form $\xi_t(q) = \exp\left(-\eta_t r_t^{\psi} q^{-\theta}\right)$ and the change of variables $Q = F_t(q)$ yields the closed-form expression reported in the main text:

$$\widetilde{\mu}_t = \frac{\widehat{\kappa}_t}{\widehat{\mu}_t}\left[1 - \left(1 + \eta_t \Lambda_t^{-1} r_t^{\psi}\right)^{-\left(1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right)}\right].$$

The key integral identity used in this step is:

$$\int_0^1 Q^x \left(-\log Q\right)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}} dQ = (1+x)^{-\left(1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right)} \Gamma\left(1 - \frac{\alpha\rho}{1-\rho\theta(1-\alpha)}\right).$$

A.7 Proof of Proposition 4

We first write $\overline{\mu}_{ut}$ into

$$\begin{aligned}
&\overline{\mu}_{ut}(Q;\Delta_i,\Delta_d) \\
&= \int_{F_t^{-1}(q)}^1 \left[1 - e^{-\widehat{\eta}_t(1-\Delta_d)^{\chi} r_t^{\psi} q^{-\theta}}\right] \left[q^{\alpha} i^{\beta} (1+\Delta_i)^{\beta} d^{\gamma} (1+\Delta_d)^{\gamma}\right]^{\rho\theta} dF_t(q) \\
&= Z_t(\Delta_i)(1+\Delta_i)^{\beta\rho\theta}(1+\Delta_d)^{\gamma\rho\theta} \int_Q^1 \left[1 - x^{\widehat{\eta}_t(1-\Delta_d)^{\chi} r_t^{\psi}}\right] (-\log x)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}} dx
\end{aligned}$$

where $Z_t(\Delta_i,\Delta_d)$ is an object that summarizes the aggregate impact of Δ_i and Δ_d. We do not need pay attention to this term because it will be canceled out in the upward citation measure. Clearly $\partial\overline{\mu}_{ut}(Q;\Delta_i,\Delta_d)/\partial\Delta_i > 0$. For Δ_d, the partial derivative (evaluated at $\Delta_i = \Delta_d = 0$) is

$$\begin{aligned}
&\frac{\partial\overline{\mu}_{ut}(Q;\Delta_i,\Delta_d)}{\partial\Delta_d}\Big|_{\Delta_{i,d}=0} \\
=&\gamma\rho\theta \int_Q^1 \left[1 - x^{\widehat{\eta}_t r_t^{\psi}}\right] (-\log x)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}} dx - \chi\widehat{\eta}_t r_t^{\psi} \int_Q^1 x^{\widehat{\eta}_t r_t^{\psi}} (-\log x)^{1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}} dx \\
\leq& 0
\end{aligned}$$

if and only if

$$\chi \geq \frac{\gamma\rho\theta}{\widehat{\eta}_t r_t^{\psi}} \frac{\int_Q^1 \left[1 - x^{\widehat{\eta}_t r_t^{\psi}}\right] (-\log x)^{-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}} dx}{\int_Q^1 x^{\widehat{\eta}_t r_t^{\psi}} (-\log x)^{1-\frac{\alpha\rho}{1-\rho\theta(1-\alpha)}} dx}$$

note the right hand side is bounded: 1) if $Q \to 0$, it is finite, and 2) if $Q \to 1$, it approaches to $\gamma\rho\theta$ by using L'Hospital's Rule twice and note $\left(1 - Q^{\widehat{\eta}_t r_t^{\psi}}\right) / (-\log Q) \to \widehat{\eta}_t r_t^{\psi}$. Therefore, we can always find a χ large enough so that the derivative is negative.

A.8 Sensitivity Analysis

A.9 Construction of Model and Data Moments

Using the merged sample between the PatentsView and Compustat firms, we construct the moments for the data as follows. We keep the firms that have at least 10 years of data. This helps us to focus on a balanced panel to minimize the firm churning. We extract the unique gvkey IDs and perform 200 bootstraps, so one draw is one firm. For each bootstrap sample, we merge back to the original sample to get the firm panel. For each sector $\times$ year in one bootstrap, we compute moments on citation flows, the ratio of software capital to knowledge capital, the dispersions of firm-level variables, etc. We then average the moments across all sectors $\times$ years. We eventually have 200 vectors of moments, each of one individual bootstrap sample. We use these moments to construct variance-covariance matrix, which is then used in simulated method of moments estimation. Finally, we average all the moments across all bootstrap samples, and we use the model to target this averaged moment.

A.10 Vertical Knowledge Diffusion

In the model, firms accumulate knowledge through experimentation (κ is the intensity) or meeting other firms (μ is the intensity). The composition of these two margins depends on the level of knowledge capital. Firms with different knowledge capital have different approaches to knowledge accumulation. Firms with low knowledge capital accumulate knowledge mainly through meetings with others. Firms with a high level of knowledge capital rely mainly on knowledge creation.

In the model, citation patterns vary systematically across the knowledge distribution: low-knowledge firms rely relatively more on learning from others, while high-knowledge firms rely relatively more on internal innovation. Comparing economies with low versus high costs of developing data and software, the model implies that both measures of knowledge diffusion

increase when data&software is more expensive, because firms invest less in nontransferable inputs and ideas become easier to learn from. This pattern is consistent with our empirical findings.

B Data Appendix

B.1 Matching Assignees of PatentsView to Compustat

We follow Dorn et al. (2020). We use PatentsView instead of USPTO because PatentsView covers more recent patents. In PatentsView, the most recent application year is 2021; given an average publication lag of around 4–5 years, we observe published patents up to 2016–2017. We restrict attention to granted patents.

The steps for patent matching are as follows:

1. We download the patent data from Compustat and the Compustat North America data from WRDS.

2. We clean assignee names in PatentsView and firm names in Compustat. This step standardizes abbreviations and removes illegal characters and punctuation.

3. We send the punctuation-free names from both sides (the patent side and the Compustat side) to the Microsoft Bing Search Engine (by using their paid API);

4. The assignee and the company name is a match if either of the following five criteria satisfies

    - The punctuation-free names match exactly.
    - The Derwent-standardized names match exactly.
    - One of the top five search results for the assignee exactly matches the firm website listed in Compustat.
    - The top five search results for the assignee and for the Compustat firm share at least two exact matches.

- The punctuation-free names share the same first word, and the firm website listed in Compustat is a partial string of one of the top five search results for the assignee (or vice-versa).

In Table Dorn et al. (2020), we list the percentage of patents that can be assigned to Compustat firms. We in addition compares the matching rate for our exercise and that of the Dorn et al. (2020)'s. As we largely follow their method, the match rate are very similar.

Table 2: Matching Rate of Patent Assignee and Firm Name

year	total patents	our's (match rate)	Autor et. al. (2020) (match rate)
2000	157595	77809(0.49)	76733(0.49)
2001	166158	82044(0.49)	78802(0.47)
2002	167400	84800(0.51)	83600(0.50)
2003	169077	86732(0.51)	85615(0.51)
2004	164384	85451(0.52)	84173(0.51)
2005	143891	75708(0.53)	74556(0.52)
2006	173822	93395(0.54)	92016(0.53)
2007	157331	84768(0.54)	83263(0.53)
2008	157788	86247(0.55)	84653(0.54)
2009	167463	92001(0.55)	90090(0.54)
2010	219848	118470(0.54)	115670(0.53)
2011	224871	120005(0.53)	115545(0.51)
2012	253633	133918(0.53)	128407(0.51)

Notes: This table shows the match rate between the PatentsView and Compustat firms.

B.2 Note on the Measurement of Software

The variable we use is the "Purchased / Capitalized software (item 1043)" under Compustat Other Intangibles (item 1040). According to Peters and Taylor (2017), intangible asset created are mostly expensed and almost never appear on asset on the balance sheet. For software,

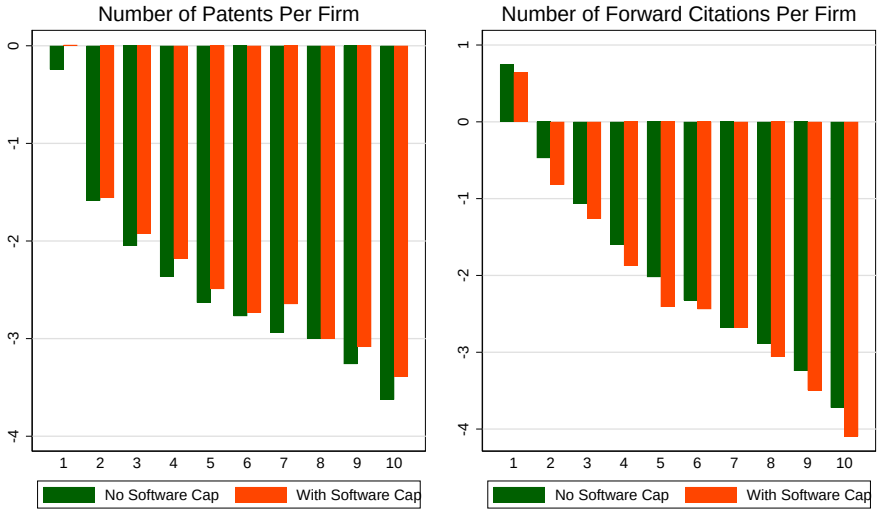
the exception is that a firm can start capitalize the spending only after the product reaches technological feasibility (for externally sold software)[3] or reaches the coding phase (for internally used software) [4].

More specifically, according to the standard accounting practice, cost of computer software developed or obtained for internal use that shall be capitalized include only the following: fees paid to third parties for services provided to develop the software during.

In our case, we drop all firms that design software for sale (i.e., NAICS 511210). The remaining sample therefore reflects capitalized costs for software developed for internal use.

B.3 Alternative Measures of Diffusion

Table 3: Heterogeneous Diffusion



Notes: This figure shows firms' patenting and diffusion patterns. The sorting of knowledge bins follows Figure 9. The left panel shows the number of patents divided by firms' knowledge capital. The right panel shows total citations divided by knowledge capital.

[3]prodct design must be complete, all resources necessary to produce must be available, the completeness of the detailed program design must have been confirmed and high-risk development issues must have been resolved.

[4]as opposed to the preliminary stage, application stage, and the post-implementation stage.

B.4 Internal Software as a Measure of Firm Secrecy

We briefly describe the accounting treatment of software and explain why our proxy reflects internally developed software and is related to firm secrecy.

In general, intangible assets are mostly expensed and almost never appear on the balance sheet Peters and Taylor (2007). For software, an exception is that firms can capitalize spending only after the product reaches technological feasibility (for externally sold software) or the coding phase (for internal-use software). The relevant Compustat item is capitalized software within "Other Intangibles." For our empirical exercise, we drop software publishers (NAICS 511210). Software capital in the remaining firms therefore mostly reflects internally developed software.[5]

This is useful for our purposes. A common view is that R&D often generates relatively transferable knowledge (e.g., product designs). By contrast, capitalized software includes compensation for programmers, consulting fees for third-party developers, testing, and direct and indirect overhead costs. The resulting proprietary software and data-processing methods are typically usable only within the firm (Haskel and Westlake, 2018; PwC, 2021). We therefore treat software stock as a proxy for firm secrecy.

B.5 Citation Flows as a Measure of Knowledge Diffusion

A common approach in the literature is to measure knowledge diffusion using patent citation links (Akcigit, Celik, and Greenwood, 2016). Since we focus on across-firm diffusion, we measure diffusion from firm B to firm A in year t using the number of citations from A to B. We focus on within-industry upward citations, which are important for business dynamism (Akcigit and Ates, 2023). We define firm B to be a "better" firm for A in year t if B's average knowledge capital over the previous three years exceeds A's.[6] We then construct citation links by aggregating citations across firms.

[5] Capitalizable costs of internal-use software include fees paid to third parties for development services (e.g., external direct costs of materials and services consumed in developing or obtaining internal-use computer software).

[6] Using a historical average makes firm rankings slow-moving and reduces the impact of firm turnover. The results are robust to using four or more years.

B.6 Summary Statistics

Table 4 reports summary statistics for the key variables used in the empirical analysis. The sample period is 2000–2015. We restrict the sample to firms that appear in every year to form a balanced panel.

For each firm, we compute time averages of the variables of interest and collapse the data to the firm level. The final sample contains 2,500 firms, of which 380 report positive software capital. This suggests that relatively few firms capitalize internal software in Compustat.

Panel A shows the distributions of sales, knowledge capital, software capital, number of patents, and within-sector knowledge flows (measured by average upward citations). Firms in our sample are large (median annual sales are around $800 million), and the distributions of sales and knowledge capital are highly skewed. Most firms have patent applications in the sample period, and patents cite on average several "better" firms (industry peers with higher knowledge capital over the previous three years). Panel B conditions on firms with positive software capital; these firms are larger on average, have higher knowledge capital, and file more patents. Their patents cite fewer better firms on average, in part because there are fewer firms above them in the industry ranking.

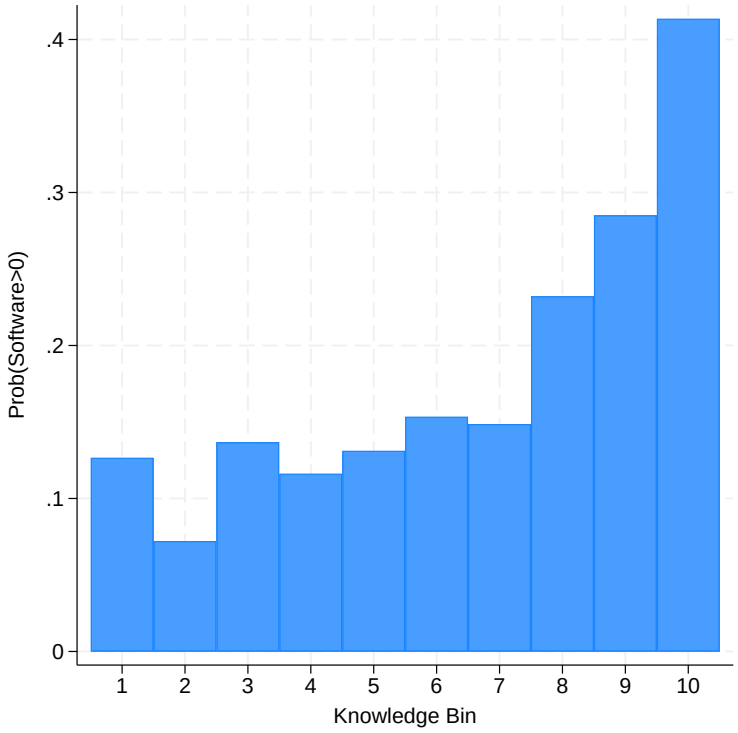
Table 4: Summary Statistics

	N	Mean	SD	P10	P25	P50	P75
Panel A: Full Sample							
Sales	2422	32982.37	106392.50	0.00	33.91	824.34	10300.08
Knowledge Capital	2111	479.53	1556.72	0.63	14.38	61.61	209.67
Software Capital	379	3623.86	18067.25	0.26	2.76	22.41	165.70
Number of Patents	2422	309.15	977.47	0.00	13.00	38.00	151.00
Within-sector Knowledge Flow	2422	720.58	1699.23	2.00	35.00	142.00	540.00
Panel B: Firms with Software Capital							
Sales	373	82863.37	170672.38	174.92	1137.31	12918.20	75683.98
Knowledge Capital	371	974.33	2305.37	2.31	24.65	141.80	759.39
Software Capital	379	3623.86	18067.25	0.26	2.76	22.41	165.70
Number of Patents	363	602.46	1324.56	16.00	33.00	122.00	486.00
Within-sector Knowledge Flow	373	1080.36	2198.23	9.00	42.00	187.00	922.00

Note: This table shows the firm-level summary statistics. Panel A shows the full sample result, and panel B focus on firms report at least one years of software capital in the sample period. The sales, knowledge capital and software capital are all in millions of USD. Number of patents measures the total patent holding counts of a firm, and within-sector knowledge flow is the total backward citation counts of a firm.

In this section, we examine the extensive margins of investment in software. Specifically, we plot the relationship between the firm's knowledge capital and the probability of reporting software expenditure in the left panel of Figure 8. We divide the firms into 10 bins based on the logarithm of the firm's knowledge capital. The y-axis is the probability of reporting software. The figure shows that the probability of reporting software increases with knowledge capital. For the top 10% companies, the probability of reporting software is 35%, while for the bottom 10% of firms, the probability is 10%.

Figure 8: Software in the Cross-Section (Extensive Margin)



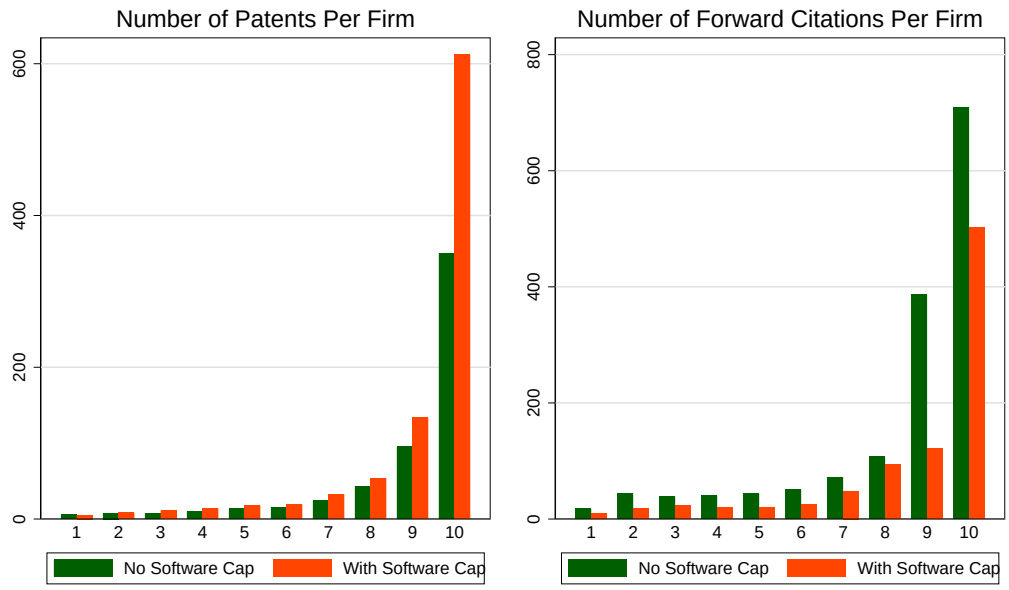
Note: The left panel of this figure plots the relationship between firm's knowledge level and the probability of reporting software expenditure. The x-axis is the 10 bins based on the logarithm of the firm's knowledge capital, and the y-axis is the probability of reporting software. The right panel plots the relationship between knowledge and software. The vertical axis is the averaged logarithm of software within each knowledge bin.

We see from the last figure that firms with more knowledge are, on average, more likely to develop software. Here, we show that firms with more knowledge capital diffuse less knowledge. We follow the same procedure as above to sort firms into 10 bins based on the logarithm of knowledge capital (each bin contains an equal number of observations). Within each bin, we compute the average number of patents and the average number of forward citations, separately for firms that report software capital and firms that do not. The result is shown in Figure 9.

The left panel is perhaps not a surprise, the firms in the top 10% publishes 30 times more patents than a median firm. In addition, firm with software stock publishes uniformly more patent compared to firms without, especially for the top firms. For example, top firms with

software publishes almost twice as much patents. However, in the right panel shows these top firms with software actually has around 30% less forward citations compared to firms does not report having software capital. Similar gaps in forward citations are also observed in all other groups of firms. This fact suggests that software investment seems to be related to a reduced (external) "quality" of patent publications.

Figure 9: More Patenting with Less Forward Citations



Note: This figure shows the patent publications and patent citations for firms with different knowledge level, with and without software capital. We regress firms knowledge stock on the industry dummy, and extract the residual. Then residual is then sorted into 10 equalsized bins. Within each bin, we compute the total patent publications and forward citation links for each firm within the bin, which are then averaged across all firms within the bin, separately for firms with/without software stock.

B.7 Sector-level Results

Table 5 presents the results from regressing sector-level knowledge diffusion on sectoral software stock. The results are consistent with the firm-level findings: the coefficient on sectoral software stock is negative and statistically significant. These sector-level patterns confirm that the negative relationship between software investment and knowledge diffusion is robust at higher levels of aggregation.

Table 5: Sector Level Regressions: Software and Knowledge Diffusion

	(1)	(2)
data and software	-0.8191	-1.0082**
	(0.1257)	(0.0232)
Concentration Controls	✗	✓
Sector-FE	✓	✓
Year-FE	✓	✓
Observations	135	135
R^2	0.2866	0.3046

Note: This table shows the sectoral results. The first column is a standard panel regression where we include sector and year fixed effect. In the second column, we include sector-year-level concentration measures. The dependent variable is the upward citations averaged within each industry-year cell, and the p-values are in parentheses. Standard errors are clustered at the sector level.