

Measuring Creative Destruction[†]

Ali Kakhbod

UC Berkeley

Leonid Kogan

MIT, NBER

Peiyao Li

UC Berkeley

Dimitris Papanikolaou

Northwestern, NBER

September 18, 2026

Abstract

We construct a firm-level measure of creative destruction from the similarity between a firm's 10-K technology description and other firms' patented innovations. Motivated by a simple model, we identify two channels of displacement, one in the product market and one through the firm's production technologies, and build a composite measure summarizing both. Higher composite displacement is associated with significant declines in firm profits, revenue, employment, and capital. We find no evidence that industry-level creative destruction raises subsequent growth, but we do see significant reallocation: industries with more creative destruction experience greater dispersion in firm growth rates.

Keywords: Innovation, Displacement, Patents, Machine learning.

JEL Classification: C1, G10, G11, O3.

[†]First Draft: August 2024. This version September 2026. We are grateful to Jan Bena, Maxime Bonelli, Amir Kermani, Martin Lettau, Dmitry Livdan, and Gordon Phillips; participants of the 2024 NBER meeting (Big data) and 2025 WFA annual meeting; and seminar participants at the University of California, Berkeley and Panagora (Crowell memorial seminar) for helpful comments. We thank Borys Babiak for outstanding research assistance. Ali Kakhbod and Peiyao Li are with the Haas School of Business at UC Berkeley. Leonid Kogan is with the Sloan School of Management at MIT and NBER. Dimitris Papanikolaou is with the Kellogg School of Management at Northwestern University and NBER. E-mails: Kakhbod (akakhbod@berkeley.edu), Kogan (lkogan@mit.edu), Li (ojhfklsjhl@berkeley.edu), Papanikolaou (d-papanikolaou@kellogg.northwestern.edu).

1 Introduction

Creative destruction—the process by which new innovations displace existing technologies and the firms that rely on them—is at the center of two long-running views of how innovation shapes the economy. The Schumpeterian growth tradition—[Aghion and Howitt \(1992\)](#); [Klette and Kortum \(2004\)](#); [Romer \(1990\)](#); [Schumpeter \(1942\)](#)—treats creative destruction as the engine of long-run productivity growth: rivals’ innovations displace older incumbents, and the net result is a higher level of aggregate output. A complementary firm-dynamics tradition—[Akcigit and Ates \(2019\)](#); [Davis and Haltiwanger \(1992\)](#); [Fernández-Villaverde, Yu and Zanetti \(2025\)](#); [Garcia-Macia, Hsieh and Klenow \(2019\)](#); [Haltiwanger \(2012\)](#); [Kogan, Papanikolaou and Stoffman \(2020\)](#), with [Acemoglu et al. \(2018\)](#) integrating the two views in a single framework— emphasizes reallocation, entry and exit, employment churn, and dispersion in firm growth as central margins of creative destruction. Empirical analysis of the creative destruction mechanism requires a measure of how exposed each firm is to the innovations being produced elsewhere in the economy. Existing proxies capture important but distinct facets of this exposure. Product-market similarity speaks to rivalry in output space ([Hoberg and Phillips, 2016](#)), patent-citation-based obsolescence tracks the declining relevance of the external patents a firm builds on (the patents it cites) ([Ma, 2025](#)), and aggregate innovation value identifies industry-level technology shocks but is silent on which firms within an industry are most affected by these shocks ([Kogan et al., 2017](#)). Patent-citation duration captures the typical length of a product’s technological life cycle at the industry level ([Bilir, 2014](#)) — predictable turnover within an established technological paradigm — but not which individual firms face displacement when innovation shifts across paradigms. We develop a granular firm-level measure of how exposed a focal firm’s operations and product mix are to the technologies embedded in other firms’ recent innovations.

We construct an annual firm-level measure of Innovation Displacement Exposure (IDE) for the universe of U.S. public firms. The measure records how relevant other firms’ recent innovations, weighted by their economic value, are to the focal firm’s existing technologies. A stylized model in [Section 2.1](#) motivates two channels of displacement: a process channel, where rivals’ advances tighten factor markets for the production methods the firm employs, and a product channel, where rivals’ advances erode the quality-adjusted prices the firm can sustain in its markets. We construct

one similarity-based measure for each channel from large-language-model summaries of patents and 10-K technology disclosures, embedded in a common vector space and weighted by aggregate patent value (Kogan et al., 2017). The panel covers U.S. public firms from 1997 to 2019. Armed with this firm-level measure, we revisit the long-running question of whether creative destruction primarily lifts the level of industry productivity or primarily reorganizes activity across firms within an industry.

Our primary specification projects forward profit growth on composite displacement and an orthogonal asymmetry component, with 3-digit SIC industry and sector-by-year fixed effects and a standard control vector. A one-standard-deviation increase in composite displacement is associated with a decline in profit growth of about 11 log points at one year and 29 log points at five years. Composite displacement also loads negatively on subsequent revenue, market share, employment, and physical and intangible capital, with magnitudes that grow with the horizon. Two forward-looking text-based risk measures—specialist labor-market tightening and product obsolescence—both rise with composite displacement, consistent with propagation through process and product margins in parallel.

At the industry level, our firm-level measure provides firm-level evidence bearing on the two views with which the paper opens. We find no evidence that composite displacement raises average industry productivity, in either the Compustat sum-of-sums series or the BEA’s published 4-digit NAICS labor-productivity index, which covers the full universe of producers. By contrast, on the reallocation margin, a 0-to-100-percentile move in industry composite displacement widens the within-industry P90–P10 spread of one-year forward log revenue-per-employee growth by 0.571. In our sample, creative destruction at the industry level manifests as reallocation, not as a level shift.

IDE carries incremental information beyond existing displacement proxies. Adding a product-market similarity measure built from our own 10-K embeddings following Hoberg and Phillips (2016) leaves the composite-IDE coefficients negative at every horizon, and the interaction of composite IDE with Product Market Similarity is informative about how quickly displacement cumulates into profits (Section 5.1). Adding the patent-citation-based technological obsolescence measure of Ma (2025) on the patenting subsample leaves the IDE coefficients essentially unchanged while the citation-based measure retains little incremental explanatory power. Because IDE is built

from 10-K text rather than patent citations, it covers all 10-K filers, including the majority of firms that do not patent.

Conceptually closest to our paper is [Kogan et al. \(2017\)](#), who measure a firm’s exposure to creative destruction as the dollar-weighted innovation output of other firms in its 3-digit SIC industry. We refine that approach by replacing industry assignment with a direct, firm-specific match between each firm’s product and technology stacks and the technologies described in other firms’ patents, and by separating displacement into a *product-market* channel and a *process* channel that the model treats as conceptually distinct—channels that share the same underlying force but propagate through output prices versus factor prices.

Two scope restrictions are worth flagging here. IDE is built from the textual content of patent filings and 10-K technology disclosures. Forms of creative destruction that do not leave traces in either source—process know-how held as a trade secret, business-model innovation that is not patented, design changes protected through copyright or trademarks rather than utility patents, and innovations originating in firms that are not publicly listed and do not file 10-Ks—are by construction outside the measure. Our results should accordingly be read as evidence on creative destruction within the patent-and-text part of the innovation distribution, not the full distribution of innovations.

The paper contributes to three literatures. First, relative to work that identifies technology shocks from Solow residuals, VARs, patent counts, or R&D spending ([Alexopoulos, 2011](#); [Kortum, 1993](#); [Kortum and Lerner, 1998](#); [Shea, 1998](#)), we provide a direct, firm-level measure of exposure to other firms’ valuable innovations. Relative to citation-based displacement measures ([Ma, 2025](#)), IDE covers all 10-K filers rather than patenting firms alone, and relative to industry-peer measures ([Kogan et al., 2017](#)), IDE matches each firm directly to the content of rivals’ innovations and separates product-market from process-side displacement. Second, by providing firm-level evidence bearing on the Schumpeterian-growth view and the reallocation view of creative destruction at the industry level, we contribute to the literature on the real effects of innovation and the balance between knowledge spillovers and business stealing ([Acemoglu et al., 2018](#); [Akçigit and Ates, 2019](#); [Davis and Haltiwanger, 1992](#); [Kogan, Papanikolaou and Stoffman, 2020](#)). The design relates to [Bloom, Schankerman and Van Reenen \(2013\)](#), who measure firm-pair proximity separately in technology space and in product-market space and find opposite-signed effects of rivals’ R&D:

positive knowledge spillovers through the first, business stealing through the second. IDE reverses the sign of the technology-space margin: because it measures the overlap of other firms’ valuable innovations with the focal firm’s own technology stack, technological proximity here captures displacement exposure rather than spillovers, while product-market similarity plays the rivalry role. The contrast between composite displacement—the level of external innovation pressure—and an orthogonal asymmetry component lets us separate the level effect from the cross-channel margin: the level loads broadly on profits and factor demand and on within-industry dispersion, while the asymmetry component loads on capital. The industry-level finding that creative destruction shows up as reallocation rather than a level shift is delivered by composite displacement alone (Section 6); it is separate from this firm-level contrast between the composite and the asymmetry component. Third, the paper joins a growing body of work using NLP and large language models in economics and finance (Fedyk et al., 2024; Gentzkow, Kelly and Taddy, 2019; Hansen, McMahon and Prat, 2018; Kelly et al., 2021); concurrent applications include Breitung and Müller (2025) on firm networks and Caskurlu, Hoberg and Phillips (2024) on innovation trends.

The remainder of the paper is organized as follows. Section 2 describes the construction of the IDE measures, beginning with the economic motivation for the process–product decomposition. Section 3 examines the properties of IDE, including cross-sectional patterns, the process–product decomposition, and time-series variation. Section 4 presents the main empirical results on firm-level outcomes. Section 5 demonstrates the incremental value of IDE relative to existing displacement measures. Section 6 examines industry-level outcomes. Section 7 presents robustness analyses. Section 8 concludes.

2 Measuring Innovation Displacement Exposure

This section develops the IDE measures. We first lay out a model of how external innovations can reduce a firm’s profits through two related channels and derive the model-implied objects that an empirical measure should capture (Section 2.1). Section 2.2 describes the data. Section 2.3 constructs the channel-specific empirical IDE measures, mapping the model’s theoretical objects to text-based representations of firms’ technologies and innovations. Section 2.4 characterizes the conditional cross-section of process and product IDE against standard firm characteristics. Section

2.5 summarizes the channel-specific pair via principal component analysis as a composite level of displacement and an orthogonal asymmetry component.

2.1 Model and model-implied IDE

We sketch a model, fully derived in Appendix D, in which external innovations enter the focal firm's profit forecast through two channels. In the *product channel*, rivals' product-quality improvements lower their quality-adjusted prices and shift expenditure away from the focal firm. In the *process channel*, external innovations close to processes used by the focal firm raise demand for specialist labor in those processes and can also improve the process productivity of innovating rivals. Thus, the product channel operates through output-market competition, while the process channel operates through the focal firm's net process-cost position relative to competitors.

By a *process*, we mean a technology, method, or capability used in production, delivery, or operation. Firms selling similar products can rely on different process stacks. For example, integrated steel mills and mini-mills both produce steel, but they rely on different furnaces, input pipelines, and specialized labor. Similarly, brick-and-mortar retailers and e-commerce retailers both sell goods, but they rely on different combinations of stores, fulfillment centers, automation, and logistics. The model treats each firm as drawing on a basket of such processes.

Demand and production.—Within each product market m , firms compete under CES demand with elasticity $\theta_m > 1$. The focal firm's revenue share depends on its quality-adjusted price relative to rivals. Each firm produces using multiple processes $j \in \mathcal{J}$, combined through a CES aggregator with elasticity $\psi > 1$. Within each process, output is Cobb–Douglas in firm-specific technology capital $k(f, j)$ and process-specific labor $\ell_{fm}(j)$, with capital share $\alpha \in (0, 1)$. Process-specific labor is inelastically supplied in the short run. Hence, external process- j innovations raise the equilibrium process- j wage.

Product channel.—When a rival g in market m improves product quality, $d \log Q_{gm} > 0$, its quality-adjusted price falls. Under fixed within-market expenditure, the focal firm's product-level profit response is

$$d \log \Pi_{fm}^{\text{prod}} = -(\theta_m - 1)\eta_m \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} d \log Q_{gm}, \quad (1)$$

where

$$\eta_m \equiv \frac{1}{1 + \alpha(\theta_m - 1)} \in (0, 1] \quad (2)$$

captures the dampening from endogenous marginal-cost movements under decreasing returns in variable labor. Aggregating across product markets with profit weights $\pi_{fm} \equiv \Pi_{fm}/\Pi_f$ gives

$$d \log \Pi_f^{\text{prod}} = - \sum_{m \in \mathcal{M}_f} \pi_{fm} (\theta_m - 1) \eta_m \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} d \log Q_{gm}. \quad (3)$$

Equation (3) holds within-market expenditure fixed. Once expenditure reallocates across markets, the sign survives under the nested ordering $1 < \nu < \theta_m$, where ν is the elasticity of substitution across markets: varieties within a market substitute more readily than markets substitute for one another. Appendix D states the exact condition, that business stealing dominate expenditure reallocation, and shows that the ordering delivers it.

Process channel.—Let $z_{j,t}$ denote external innovation pressure in process j , defined in Appendix D.6 as the log change in other firms' effective process- j technology capital. Short-run labor-market clearing implies

$$\frac{\partial \log w(j)}{\partial z_{j,t}} = \kappa_{j,t} = \alpha \frac{K_{j,t}^{(-f), \text{eff}}}{K_{j,t}^{\text{eff}}} > 0, \quad (4)$$

where $K_{j,t}^{(-f), \text{eff}}$ is the external component of effective process- j technology capital and $K_{j,t}^{\text{eff}}$ is total effective process- j technology capital.

Let $s_{f,j}$ be the focal firm's process exposure share, defined as the share of variable production cost accounted for by process- j labor. Because innovating rivals may also receive direct process-capital gains, define

$$e_{g,j,t} \equiv \frac{\partial \log k(g, j)}{\partial z_{j,t}}, \quad \rho \equiv \frac{\alpha}{1 - \alpha}. \quad (5)$$

The net process-cost loading of firm g is

$$b_{g,j,t} \equiv s_{g,j} (\kappa_{j,t} - \rho e_{g,j,t}). \quad (6)$$

For the focal firm, $e_{f,j,t} = 0$, so

$$b_{f,j,t} = \kappa_{j,t} s_{f,j}. \quad (7)$$

For innovating rivals, $b_{g,j,t}$ can be smaller, or even negative, if their process-capital gains dominate the wage-pressure component. Define the market-average net process-cost loading as

$$\bar{b}_{m,j,t} \equiv \sum_{g \in \mathcal{F}_m} \sigma_{gm} b_{g,j,t}. \quad (8)$$

The product-level process-channel profit response is

$$\frac{\partial \log \Pi_{fm}^{\text{proc}}}{\partial z_{j,t}} = -\frac{(1-\alpha)\Lambda_m}{1+\alpha\Lambda_m} \bar{b}_{m,j,t} - (\theta_m - 1)\lambda_m (b_{f,j,t} - \bar{b}_{m,j,t}), \quad (9)$$

where

$$\lambda_m \equiv \frac{1-\alpha}{1-\alpha+\alpha\theta_m} \in (0,1) \quad (10)$$

is the pass-through dampening term, and $\Lambda_m \geq 0$ captures the sensitivity of market expenditure to the market price index. The first term is the market-expenditure component. The second term is the within-market relative-cost component. Under fixed market expenditure, $\Lambda_m = 0$, so the process channel operates only through the focal firm's net loading relative to the market average:

$$b_{f,j,t} - \bar{b}_{m,j,t}. \quad (11)$$

Aggregating across product markets and processes gives

$$d \log \Pi_f^{\text{proc}} = \sum_{j \in \mathcal{J}} \left[\sum_{m \in \mathcal{M}_f} \pi_{fm} \frac{\partial \log \Pi_{fm}^{\text{proc}}}{\partial z_{j,t}} \right] z_{j,t}. \quad (12)$$

Implications for measurement.—The model motivates two empirical exposure measures. Product IDE aggregates valuable external innovations by their proximity to the focal firm's product technologies. Process IDE aggregates valuable external innovations by their proximity to the focal firm's internal production and operating technologies. Total profit displacement is the sum of the two channels:

$$d \log \Pi_f = d \log \Pi_f^{\text{prod}} + d \log \Pi_f^{\text{proc}}. \quad (13)$$

The empirical IDE measures capture the focal-firm exposure component of the model. The full structural mapping also depends on market shares, pass-through, process-wage responses, rival

process-capital gains, and market-expenditure responses, which are not separately observed. We therefore interpret the estimated coefficients as reduced-form associations between focal-firm exposure to valuable external innovations and subsequent profit growth. Section 2.5 summarizes the joint variation in $(IDE^{\text{process}}, IDE^{\text{product}})$ using a composite displacement component and an orthogonal asymmetry component.

2.2 Data and sample

We combine patent data from the USPTO, aggregate innovation values from Kogan et al. (2017), and firm characteristics from COMPUSTAT (profits, R&D expenses, capital expenditures, employment, total assets, capital stock, market capitalization, and 3-digit SIC industry codes). These datasets are merged using the PERMNO firm identifier and fiscal year. We then merge firm data with innovation and technology embeddings. A detailed description of the data construction procedure is in Appendix A.

Our final dataset spans 1997 to 2019. Across the full firm-year panel, 70,236 firm-years carry valid IDE scores, and the one-year profit-growth regressions in Section 4.1 use 50,028 of them. Our technology data are extracted from 10-Ks downloaded from EDGAR, which begin in 1993; however, our displacement measures use a five-year look-back window, so the earliest year for which we can compute IDE is 1997. The firm panel excludes financials (SIC 6000–6799) and utilities (SIC 4900–4949), a screen applied when the 10-K filings are collected, so 28 of the Fama–French 30 sectors appear in our sample. The average one-year profit growth in the sample is 4.8 log points and the average five-year profit growth is 18.1 log points.

2.3 Empirical construction

We now describe how we map the theoretical objects summarized in Equations (3) and (12) to data. The construction builds on two text-based representations of each firm: an *innovation stack* summarizing the firm’s recent patents, and a *technology stack* summarizing its existing technologies from 10-K filings. We first use large language models to produce structured summaries of these documents, then embed the summaries as numerical vectors using a text embedding model, and finally compute pairwise cosine similarities between each firm’s technology embeddings and every

other firm’s innovation embeddings, weighted by the economic value of the innovations.

Innovation stack. We define firm g ’s innovations in year t as the patents it receives over the window $[t - 4, t]$. For each year in this window, we use an LLM (GPT-4o-mini) to summarize the firm’s patent abstracts into a structured technology description. The LLM input is the patent abstract; we run GPT-4o-mini at temperature zero (we do not pin a seed because OpenAI’s seed parameter is in beta and does not guarantee determinism). Only granted patents enter. For firm-years whose concatenated abstracts exceed 120,000 tokens (fewer than 1 percent of firm-years), we draw a random subset of 500 abstracts before summarizing. The prompt instructs the model to identify general themes and common topics across the firm’s patents, provide individual summaries for patents that do not fit common themes, and avoid speculation beyond what is explicitly stated in the abstracts (the full prompt is in Appendix A.8; Appendix A.8 shows that alternative prompt formulations produce nearly identical summaries, with average cosine similarities above 90%). We concatenate the annual summaries chronologically over the five-year window and embed the concatenated text as a single vector $\mathbf{Innov}_{g,t}$ using a text embedding model (text-embedding-3-large, producing 3072-dimensional embeddings).

Technology stack. We characterize firm f ’s existing technology base using its 10-K filings over the same five-year window $[t - 4, t]$. We rely on 10-Ks because they provide a comprehensive, standardized description of firms’ operations, strategies, risks, and investments, including detailed discussion of the technologies used in production and in delivering products and services. At the same time, the length and heterogeneity of these filings make direct extraction challenging at scale. We therefore use an LLM-based pipeline to convert each 10-K into structured technology summaries that retain only the technologically salient content.¹

We implement summarization in two stages to balance extraction quality and cost. In the first stage, a cost-efficient model (GPT-4o-mini) generates an overall technology summary from the full 10-K text. We pass plain text from the start of the 10-K through Item 7 (Management’s Discussion and Analysis), with footers and embedded scripts stripped but numerical values preserved; exhibits are excluded to fit the 128,000-token GPT-4o-mini context window. The first

¹Many papers in finance and economics have used 10-K disclosures to extract information related to firm performance and strategy. We provide a partial list and discussion in Appendix A.4.

stage runs at temperature zero; the second stage runs at the model’s default sampling settings.² In the second stage, because separating technologies used in internal operations from those embodied in customer-facing outputs is a linguistically demanding task, a more capable model (GPT-5.1) refines the overall summary into two targeted summaries: a *process-technology summary*—covering technologies used internally for operations, production, and R&D (here “R&D” denotes the tools and systems the firm currently uses to support research—present capabilities, not the firm’s forward-looking research program, which the patent-based innovation side captures)—and a *product-technology summary*—covering technologies offered as products and services. In the prompt itself we describe the process category to the LLM as “input-only” technologies, because that phrasing more reliably elicits the internal-versus-external distinction we want; we use the more descriptive *process technology* label in the paper. Each summary is capped at 2,000 tokens, comparable in length to the patent-based innovation summaries (the full prompts and pipeline details are in Appendix A.5; Appendix A.6 shows an example technology summary). We then embed the aggregated process and product technology summaries over the five-year window using the same text embedding model, yielding two vectors per firm-year: $\text{Tech}_{f,t}^{\text{process}}$ and $\text{Tech}_{f,t}^{\text{product}}$.

Process vs product displacement. The separation of the technology stack into process and product components is a key design choice, motivated by the model’s prediction that the two channels operate through different mechanisms (Section 2.1). Both channel measures use the same set of innovators and the same value weights $A_{g,t}$; the process–product distinction enters only through the focal firm’s technology-stack embedding. The decomposition therefore operationalizes the model’s two channels through focal-firm exposure rather than implementing them structurally, consistent with the reduced-form reading at the end of Section 2.1.

Insert Tables 1 and 2 and Figure 1 here.

Tables 1 and 2 provide intuitive evidence that this decomposition captures distinct dimensions of the firm. For Microsoft, the process-only tokens describe internal operations and research (*research, internal, production, facility, manufacture*), while the product-only tokens name customer-

²The pipeline pins model versions: gpt-4o-mini-2024-07-18 in the first stage, gpt-5.1-2025-11-13 in the second, and text-embedding-3-large for the embeddings.

facing offerings (*excel, email, live, dynamic, communication*). For Coca-Cola, the process-only tokens capture upstream production (*production, canners, raw, material, control*), while the product-only tokens are brands and product lines (*diet, classic, dasani, monster, energy*). Terms common to both summaries are correspondingly generic to the firm rather than to either channel: *microsoft, software* and *customer* in the first case, *beverage, distribution* and *drink* in the second. Figure 1 quantifies this separation over time: the fraction of lemmatized tokens appearing in both summaries is consistently small—typically at or below 20% for Microsoft and below 10% for Coca-Cola—confirming that the two summaries contain largely non-overlapping information.

IDE construction. For each firm f in year t , we compute:

$$\begin{aligned} IDE_{f,t}^{\text{process}} &= \sum_{g \neq f}^{n_{t,\text{Innov}}} \cos(\text{Innov}_{g,t}, \text{Tech}_{f,t}^{\text{process}}) \cdot A_{g,t}, \\ IDE_{f,t}^{\text{product}} &= \sum_{g \neq f}^{n_{t,\text{Innov}}} \cos(\text{Innov}_{g,t}, \text{Tech}_{f,t}^{\text{product}}) \cdot A_{g,t}, \end{aligned} \quad (14)$$

where $n_{t,\text{Innov}}$ is the number of innovative firms in year t and $A_{g,t}$ is the aggregate innovation value (relative to size) of firm g in year t from Kogan et al. (2017). Here $A_{g,t}$ is the firm-level dollar value of innovations from Kogan et al. (2017) divided by Compustat total assets in year t , winsorized at the 1st and 99th percentiles.³ Product IDE approximates Equation (3): the cosine similarity between firm g 's innovation embedding and firm f 's product-technology embedding—the embedding analog of the uncentred correlation between firms' technology-position vectors introduced by Jaffe (1986), with text embeddings in place of patent-class shares—replaces the product of the market share σ_{gm} and quality improvement $d \log Q_{gm}$, while the innovation value $A_{g,t}$ captures the economic magnitude of the innovation. Process IDE approximates Equation (12): the cosine similarity to the process-technology embedding replaces the external process shock $z_{j,t}$ from Equation (12), again weighted by $A_{g,t}$.

$IDE_{f,t}^{\text{product}}$ is high when economically important innovators introduce advances close to the focal firm's products and services—innovations that threaten its competitive position in the product

³Notation: f indexes the focal firm (previously i) and g other innovating firms (previously j); $A_{g,t}$ is firm g 's value weight and $A_{f,t}$ the focal firm's own innovation value (Section 2.4); and γ_n^k denotes the 3-digit SIC industry fixed effects in the regressions (freeing g for the other-firm index).

market. $\text{IDE}_{f,t}^{\text{process}}$ is high when economically important innovators introduce advances close to the technologies the firm relies on to run its internal operations—innovations that can substitute for, automate, or render obsolete the firm’s production capabilities. The two measures are designed to capture economically distinct channels: $\text{IDE}^{\text{product}}$ captures product-market displacement through demand erosion, while $\text{IDE}^{\text{process}}$ captures factor-market displacement through cost increases.

2.4 Firm characteristics and displacement exposure

We next relate our displacement exposure measures $\text{IDE}^{\text{process}}$ and $\text{IDE}^{\text{product}}$ to firm characteristics. We focus on the following variables: log gross profits, defined as the (log) difference between revenue (Compustat: sale) and costs (Compustat: cogs); log of the firm’s capital stock (Compustat: ppeg); log employment (Compustat: emp); profitability, defined as the ratio of gross profits to book assets (Compustat: at); the variable cost ratio (Compustat: cogs/sale); the KPSS value of firm innovation relative to book assets, $A_{f,t}$; firm age (years since IPO); and two variables summarizing the level of product market crowdedness and the revenue growth of competitor firms.

The first variable summarizes how crowded the focal firm’s product space is, following [Hoberg and Phillips \(2016\)](#). We measure crowdedness using pairwise text-based similarities between product-market descriptions. Let $\mathbf{p}_{f,t}$ denote the embedding vector of firm f ’s product-market text in year t , constructed from 10-K disclosures using the same embedding pipeline as for IDE (details in [Appendix A.5](#)), and let $\mathcal{F}_t$ denote the set of firms with a valid product-market embedding in year t . For each pair of firms f and g , define the cosine similarity between their product-market embeddings as

$$s_{fg,t} \equiv \frac{\mathbf{p}_{f,t}^\top \mathbf{p}_{g,t}}{\|\mathbf{p}_{f,t}\| \|\mathbf{p}_{g,t}\|}. \quad (15)$$

The *raw* product-market similarity sum is then

$$\text{PMS}_{f,t} \equiv \sum_{g \in \mathcal{F}_t \setminus \{f\}} s_{fg,t}. \quad (16)$$

Higher values of $\text{PMS}_{f,t}$ indicate that the focal firm sits in a crowded product space—many close peers in text-based product space—rather than that product rivalry mechanically lowers prices by itself. In all regressions, we use $\log \text{PMS}_{f,t}$, standardized over the full sample and then winsorized

at the 1 percent level on each tail. The summation universe is the same set of 10-K filers used for IDE, with the same embedding model, text span, and vector normalization, and the measure is used in Table 3 and in the interaction regressions of Section 5.1.

The second variable consists of the average sales growth of all firms, weighted by their product market similarity,

$$\text{PWSG}_{f,t} \equiv \frac{\sum_{g \in \mathcal{F}_t \setminus \{f\}} s_{fg,t} [\log \text{sale}_{g,t} - \log \text{sale}_{g,t-1}]}{\sum_{g \in \mathcal{F}_t \setminus \{f\}} s_{fg,t}}. \quad (17)$$

High values of $\text{PWSG}_{f,t}$ indicate that other firms with similar products experience higher growth in sales, which can be interpreted as a proxy for product demand.

We next ask which firm characteristics correlate with $\text{IDE}^{\text{process}}$ and $\text{IDE}^{\text{product}}$. We residualize each standardized channel-specific measure and each firm characteristic on 3-digit SIC industry and Fama–French 30 sector-by-year fixed effects, and regress the residualized measures on the residualized characteristics. By the Frisch–Waugh–Lovell theorem the coefficients are identical to those from a regression that includes the fixed effects directly; the reported R^2 , however, is a share of residualized variation, so the conditional correlations isolate within-sector-year, within-industry variation. Table 3 reports both raw correlations and multivariate regression coefficients.

Insert Table 3 here.

Examining Table 3, we see that one of the strongest predictors of process and product IDE is product-weighted sales growth—the similarity-weighted average of sales growth among a firm’s closest product-market peers, a proxy for demand conditions in the firm’s product niche. Firms whose close competitors are experiencing faster sales growth have substantially lower IDE scores than other firms, both unconditionally and once we condition on the other characteristics. These estimates suggest that firms operating in product markets with robust demand face less displacement pressure, while firms in stagnating or declining niches are more vulnerable to external innovations that threaten their competitive position. Among the other predictors, the ones that are statistically different from zero in both columns of the multivariate regression are firm age, firm size (capital stock), current profitability, and the level of product market crowdedness ($\text{PMS}_{f,t}$).

Variable cost enters significantly for product IDE alone, and log employment for process IDE alone, with a negative sign: conditional on the capital stock, more labor-intensive firms are somewhat less exposed. Conditional on the other characteristics, larger, older, and more profitable firms in more crowded product spaces are slightly more exposed to creative destruction along both measures, even though the corresponding raw correlations carry the opposite sign.

In total, the observable firm characteristics in Table 3 explain about 12 percent of the within-sector-year, within-industry variation in process and product IDE—a non-trivial share, but one that leaves the majority of the variation unexplained. The substantial residual variation indicates that process and product IDE retain information about displacement exposure beyond what standard firm characteristics pick up.

2.5 Creating a composite displacement exposure measure

In practice, the two channel-specific measures comove tightly. Across the 70,236 firm-years for which both are defined, the raw Pearson correlation between $IDE_{f,t}^{\text{process}}$ and $IDE_{f,t}^{\text{product}}$ is 0.99, but that comovement reflects primarily common variation across fiscal years, with only a small contribution from differences across coarse industries. To capture the residual variation, we demean IDE^{process} and IDE^{product} within each Fama–French 30 industry–year cell. After doing so, the within-cell Pearson correlation falls to about 0.82 (Spearman about 0.77), and a principal-axis decomposition of within-cell residuals attributes roughly 91 percent of their variance to the symmetric (joint-up) direction and 9 percent to the orthogonal tilt. Our main firm-level specification absorbs more variation than this benchmark—it uses 3-digit SIC industry and Fama–French 30 sector-by-year fixed effects—so the within-cell figures here are an upper bound on residual comovement under our regression design. Some residual shared variation likely reflects mechanical overlap: the process and product technology summaries are extracted from the same 10-K text and can recover overlapping themes. Some reflects the economic content Section 2.1 emphasizes: a firm in technology-intensive product markets typically uses technologies subject to similar external innovation pressure.

We summarize the firm’s displacement exposure by applying principal component analysis to the standardized $(IDE_{f,t}^{\text{process}}, IDE_{f,t}^{\text{product}})$ over the full firm-year sample and defining our two

summary measures as the standardized first and second principal components:

$$\overline{\text{IDE}}_{f,t} \equiv \widetilde{\text{PC}}_{1,f,t}, \quad \text{IDE}_{f,t}^{\text{tilt}} \equiv \widetilde{\text{PC}}_{2,f,t}. \quad (18)$$

We sign $\overline{\text{IDE}}$ so that its loadings on $\text{IDE}^{\text{process}}$ and $\text{IDE}^{\text{product}}$ are jointly positive (here, equal at 0.71), so that higher $\overline{\text{IDE}}$ unambiguously means more displacement on *both* sides; we sign IDE^{tilt} so that its loading on $\text{IDE}^{\text{process}}$ is positive (and on $\text{IDE}^{\text{product}}$ negative), so that positive IDE^{tilt} means process-side displacement exceeds product-side displacement. Both components are standardized to mean zero and unit variance over the full firm-year sample, and the two are orthogonal by construction. Applied to the full standardized pair, the first principal component accounts for 99.5 percent of their joint variance and the second for the remaining 0.5 percent (equal loadings up to sign, as reported above). Those shares summarize cross-firm dispersion in levels; within sector-year cells, the two channel-specific series are less tightly collinear, so IDE^{tilt} carries more independent variation for identification than the raw 0.5 percent figure alone would suggest. We carry both $\overline{\text{IDE}}$ and IDE^{tilt} into the firm-level regressions: $\overline{\text{IDE}}$ as the composite *level* of displacement and IDE^{tilt} as an orthogonal summary of the process-product *tilt*. They are distinct empirical objects with distinct economic content.

3 Properties of IDE

Armed with a direct measure of creative destruction, we now document a set of stylized facts about its distribution across industries, over time, and across firms. We find that most of the dispersion in composite IDE is *within* broad Fama–French 30 industries rather than between them; that average exposure nonetheless varies in sensible ways across those sectors, with higher means in patent- and technology-intensive activities; that the revenue-weighted aggregate moves substantially over time with its movements dominated by within-firm variation rather than compositional shifts; and that individual case studies illustrate the within-sector heterogeneity.

3.1 Heterogeneity in displacement exposure

Before turning to cross-industry patterns, we summarize the unconditional distribution of composite IDE. Table 4 reports summary statistics for $\overline{\text{IDE}}$ and the asymmetry component IDE^{tilt} , both standardized over the full firm-year sample. Composite IDE is *right-skewed* (skewness ≈ 1.9): the bulk of firm-years sit at low-to-moderate exposure—the 25th and 50th percentiles are -0.67 and -0.24 standard deviations, and the distribution is bounded below near -1.1 —while a minority of firm-years carry very high exposure, with a 99th percentile of $+3.7$ standard deviations. Displacement exposure is therefore concentrated among a relatively small set of highly exposed firms rather than spread symmetrically, consistent with the case studies below. The asymmetry component is far more dispersed in the tails (from -7.2 to $+11.1$ standard deviations) but roughly symmetric (skewness ≈ 0.15), reflecting a small number of firms whose process and product exposures diverge sharply.

Insert Table 4 and Figure 2 here.

Figure 2 plots average composite IDE ($\overline{\text{IDE}}$)—the standardized first principal component of process and product IDE—by Fama–French 30 industry, together with the within-industry interquartile range (IQR). To neutralize sample-composition effects from industries that experience heavy entry or exit during the sample period, both statistics are computed with year-equal weighting: for each industry-year we compute the cross-firm mean and the 25th and 75th percentiles, then average those statistics across years. A salient feature is how compressed the cross-industry mean differences are. Industry means span only about 0.34 standard deviations of the unconditional distribution from the lowest (Precious Metals and Mining, -0.25) to the highest (Healthcare, Medical Equipment, Pharmaceutical Products, $+0.10$). The middle 50% of firms inside a typical broad industry already spans about 0.23 standard deviations in any given year, and the most heterogeneous sectors—Healthcare, Medical Equipment, and Pharmaceuticals; Construction and Construction Materials; Recreation; and Precious Metals, Non-Metallic, and Industrial Metal Mining—reach within-industry IQRs of 0.25–0.28 standard deviations. Within-industry distributions therefore overlap almost entirely across the cross section: even in the highest-mean sectors, 25th-percentile firms lie at or below zero. A formal decomposition of firm-year variance confirms the visual pattern. Regressing composite

IDE on Fama–French 30 industry fixed effects yields an R^2 of about 0.015, so roughly 99 percent of the firm-year variance in composite IDE lies within these broad sectors. This is a different cut of the data than the within-cell decomposition of the channel-specific pair in Section 2.5, which splits the joint within-FF30 variance of $(IDE^{\text{process}}, IDE^{\text{product}})$ into a symmetric direction and an orthogonal tilt. Both decompositions point in the same practical direction: almost none of the firm-year variance in composite IDE lies between broad sectors, and what the firm-level regressions exploit is the dispersion that survives sector-by-year and industry fixed effects—about one-fifth of a pooled standard deviation, leaving substantial within-sector variation for firm-level regressions to exploit through 3-digit SIC industry and sector-by-year fixed effects.

The *level* pattern across broad sectors is nonetheless intuitive. Because composite IDE is constructed to have mean zero and unit variance in the full sample, industry means can be read as standard deviations relative to the unconditional distribution. The figure sorts industries in descending order of mean exposure (highest at the top). At the bottom are extractive and stable consumer-service sectors where patent-based displacement among listed firms is more limited, including Precious Metals and Mining, Coal, Restaurants and Hotels, Tobacco, Retail, and Transportation. At the top are technology-intensive manufacturing and product-development sectors, including Business Equipment, Healthcare, Medical Equipment, and Pharmaceuticals, Fabricated Products and Machinery, Electrical Equipment, and Chemicals. The middle of the cross section—Construction, Consumer Goods, Steel Works, Textiles, Automobiles and Trucks, Business Supplies and Shipping, and Aircraft and Railroad—mixes technology-intensive and more stable industries. The ranking is consistent with the view that incumbents in patent-intensive technology areas face the greatest external displacement pressure on average; the substantial overlap of within-industry distributions, however, cautions against reading broad sector membership as a sufficient statistic for any individual firm’s displacement exposure.

Insert Figure 3 here.

Figure 3 shows the distribution of firms across sectors. The largest sectors are Personal and Business Services, Healthcare and Pharmaceutical Products, and Business Equipment, with roughly 1,100 to 1,600 unique firms each. The smallest are Tobacco Products, Beer and Liquor, Coal, and

Textiles, each with fewer than 40.

3.2 Time-series variation

Insert Figure 4 here.

We next examine the time-series properties of composite IDE. Composite IDE is persistent at short horizons: the firm-level autocorrelation is 0.72 at one year, 0.37 at two and 0.18 at three, and is essentially zero by five. Within sector-by-year cells — the variation the specifications identify from — persistence is higher and decays more slowly, at 0.82, 0.63 and 0.48 over the same horizons. The solid line in Figure 4 plots the cumulative change in the revenue-weighted cross-sectional mean of standardized composite IDE relative to 1997—the first year for which IDE is computable, since the construction uses a five-year patent and 10-K look-back window and our 10-K technology summaries begin in 1993. The index is rebased so that 1997 is the baseline for the cumulative change. The revenue-weighted annual mean itself is volatile: it spikes around the turn of the millennium, declines through the late 2000s and early 2010s, and shows renewed strength in the mid- to late 2010s, so the cumulative path is wave-like rather than a steady upward trend in measured displacement pressure.

That said, aggregate movements could in principle reflect compositional shifts rather than genuine changes in displacement exposure among continuing firms. The remaining lines in the figure apply the [Griliches and Regev \(1995\)](#) decomposition to the cumulative change in the revenue-weighted mean, splitting it into a *within* component (changes in continuing firms' IDE holding revenue shares fixed on average), a *between* component (reallocation of revenue toward or away from high-IDE continuing firms), and *net entry/exit* (firms entering or leaving the sample). The within component tracks the aggregate closely throughout the period, reallocation between continuing firms contributes essentially nothing, and net entry and exit is small relative to the swings in the series; entering firms have lower exposure than incumbents in the year they enter, which is why the net entry/exit line drifts down. Thus the swings in aggregate exposure reflect changes in the measured exposure of continuing firms, not turnover in the sample.

The aggregate series is therefore informative about a margin distinct from the decline in U.S. business dynamism documented for firm entry rates ([Akcigit and Ates, 2019](#); [Haltiwanger, 2012](#)):

entry measures speak to new-firm formation, whereas our series tracks how exposed continuing incumbents are to external patent-based innovation given the technologies they deploy. The two need not move together, and the within-firm movements documented above are consistent with incumbents repeatedly repositioning relative to an evolving innovation frontier. The late-1990s spike, the post-crisis trough, and the partial recovery in the 2010s line up in timing with familiar technology waves, but we do not tie the series to a single structural driver.

3.3 Case studies

Next, we provide concrete examples where the displacement exposure implied by composite IDE aligns with documented displacement experiences. We also examine firms that are widely seen as *creating* displacement, where we would expect low composite IDE. Percentiles compare each firm's full-sample average of composite IDE to the distribution of firm-level averages in the same Fama–French 30 industry; sub-period percentiles compare window-specific firm means to the analogous distribution in the same window.

Boston Scientific. The healthcare sector is patent-intensive in our sample. Boston Scientific ranks at the 81st percentile of within-sector firm means on composite IDE, with an average of +0.35 standard deviations versus a sector average of firm means near +0.03. This lines up with concentrated exposure to innovation in interventional cardiology—drug-eluting stents, rhythm management, and structural heart—where generational turnover in competitor portfolios raises displacement pressure on both products and manufacturing capabilities.

3M. 3M ranks at the 71st percentile within the Fama–French “Everything Else” industry (+0.30 versus a sector average near zero), reflecting broad technology exposure across adhesives, abrasives, optical films, filtration, and healthcare-adjacent products, where patent-based advances by rivals map onto many of 3M's production processes simultaneously.

Procter & Gamble. P&G sits at the 67th percentile in consumer goods (+0.19 versus a sector mean of firm means of +0.10). Its within-sector percentile is higher in several multi-year sub-windows than in the pooled sample, running between the 80th and 96th percentile in 1997–2000, 2005–2008,

2013–2016, and 2017–2019, indicating persistently high *relative* displacement exposure throughout our sample even as sector-wide levels move over time.

Eastman Kodak. Kodak is the canonical creative destruction case: a dominant incumbent whose core technology was rendered obsolete by external innovation. In a full-sample average sense, Kodak’s composite IDE is close to the consumer-goods sector mean, but sub-period percentiles are informative. Kodak is near the top of consumer goods in 1997–2000 (about the 83rd percentile) and remains in the top quintile over 2005–2012 (including roughly the 83rd percentile in 2005–2008 and the high 80s in 2009–2012, the years leading to its 2012 bankruptcy), consistent with digital imaging innovations displacing film-era production assets. Relative exposure falls to the sector median in 2013–2016 (51st percentile), just after the reorganization, before returning to the 81st percentile in 2017–2019.

Ford Motor Co. Ford’s full-sample composite IDE is near the sector median (49th percentile), masking pronounced dynamics in the EV transition. Ford’s within-sector percentile on composite IDE rises from roughly the 38th in 1997–2000 to roughly the 68th in 2017–2019, with the increase concentrated after the late 2000s as battery, power-electronics, and software-intensive architectures become central to competition. Beginning in 2015, Ford committed \$4.5 billion to electrify 40% of its product lineup and hired over 120 new electrified-powertrain engineers, investments necessitated by external innovations in battery technology, power electronics, and autonomous-driving software; the company subsequently faced an estimated \$8 billion cost deficit against EV-native competitors.

Cisco Systems. Cisco ranks near the business-equipment median on composite IDE (45th percentile), with a firm mean slightly below the sector average of firm means. Sub-period percentiles start near the 48th in 1997–2000, fall to about the 23rd in 2005–2008, recover partially through 2013–2016 (about the 38th), and fall back to about the 21st by 2017–2019. This pattern is consistent with cloud computing, software-defined networking (SDN), and network virtualization disrupting Cisco’s hardware-centric model—during our sample period, Cisco’s Ethernet switching market share declined from approximately 62 percent in 2013 to 55 percent by 2017, and its router market share fell from 49 percent to 44 percent, as competitors like Arista Networks and

Huawei offered software-based alternatives—followed by adaptation toward software-defined and subscription-based offerings.

Apple and Tesla. The preceding cases examine firms facing displacement. An equally important validation is that firms widely recognized as *creating* displacement should have low composite IDE. Apple ranks at about the 35th percentile among business-equipment firms on average, with sub-period percentiles between roughly the 12th and the 38th, despite operating in a high-exposure sector. Apple’s proprietary ecosystem—integrated hardware, software, and services controlled through a closed platform—insulates it from external patent-based innovation relative to peers; Apple is the firm *producing* displacement for others, not experiencing it.

Tesla presents a parallel case in automobiles. When Tesla appears in Compustat from 2010 onward, it remains below the automobile-sector median on average (16th percentile for its full-sample mean). Within-sector percentiles reach only about the 18th in 2013–2016 and the 39th in 2017–2019, still well below Ford in that late-2010s window (Ford near the 68th percentile). The Ford–Tesla contrast within the same sector illustrates the asymmetry central to creative destruction: the incumbent’s vulnerability rises as the technology frontier shifts toward domains where entrants have led patenting.

In sum, these case studies illustrate four features of composite IDE. First, the measure lines up with salient displacement episodes (Boston Scientific in stents, Kodak in digital photography, Ford in EVs, Cisco in cloud networking). Second, it separates disruptors from displaced incumbents within the same industry (Apple and Tesla versus sector medians; Ford versus Tesla in the late 2010s). Third, within-industry heterogeneity is substantial even after aggregating process and product exposure: 3M towers above typical diversified manufacturers, while Apple remains in the bottom half of business equipment. Fourth, multi-year windows reveal dynamics that full-sample averages can mask (Ford, Kodak, Cisco, P&G).

4 IDE and firm performance

We now examine the predictive content of IDE for firm-level outcomes. Figure 2 documents that almost none of the firm-year variance in composite IDE lies between broad sectors; the regressions

below exploit the dispersion that survives sector-by-year and industry fixed effects, about one-fifth of a pooled standard deviation. Our primary test is whether IDE is associated with subsequent profit growth, the outcome most directly tied to the model. We then examine broader growth outcomes and examine forward-looking proxies for the process and product channels.

4.1 IDE and profit growth

We estimate the following regressions:

$$\log \Pi_{f,t+k} - \log \Pi_{f,t} = \beta_1^k \overline{\text{IDE}}_{f,t} + \beta_2^k \text{IDE}_{f,t}^{\text{tilt}} + \boldsymbol{\lambda}^{k\top} \mathbf{X}_{f,t} + \delta_{s \times t}^k + \gamma_n^k + \epsilon_{f,t}^k, \quad (19)$$

where the dependent variable is the change in log profits for horizons $k \in \{1, \dots, 5\}$. The composite displacement measure $\overline{\text{IDE}}_{f,t}$ is the first principal component of the channel-specific $\text{IDE}^{\text{process}}$ and $\text{IDE}^{\text{product}}$ measures (constructed in Section 2.3); the contrast $\text{IDE}_{f,t}^{\text{tilt}}$ is the orthogonal second component. The control vector $\mathbf{X}_{f,t}$ includes log current profit, log employment, log capital stock, profitability, product-weighted sales growth $\text{PWSG}_{f,t}$, and firm-level innovation value deflated by total assets, all defined in Section 2.4.

We construct $\text{PWSG}_{f,t}$ as in Section 2.4 (Equation (17)), using the cosine similarities $s_{fg,t}$ from Equation (15) as weights on the contemporaneous log sales growth of all other firms in the 10-K-filer universe. Sales are CPI-deflated, year shocks are absorbed by sector-by-year fixed effects, and we standardize $\text{PWSG}_{f,t}$ over the full sample, then winsorize it at the 1st and 99th percentiles on each tail.

We include 3-digit SIC industry fixed effects γ_n^k and sector-by-year fixed effects $\delta_{s \times t}^k$. The two IDE components, product-weighted sales growth, and firm-level innovation value are standardized to unit standard deviation across the entire sample. Standard errors are clustered by firm.

Before turning to the estimates, consider what Equation (19) delivers when the two channels replace the composite. Table 5 reports that regression: each channel alone is negatively and significantly associated with subsequent profit growth at every horizon, with process running from -0.101 at one year to -0.272 at five and product from -0.087 to -0.242 . Entered together both remain individually significant, but their standard errors rise by about half.

Insert Tables 5 and 6 here.

Table 6 reports the estimates. The coefficient on composite IDE is negative at every horizon. A one-standard-deviation increase in composite IDE is associated with a 10.6 log point lower profit growth rate over the next year and a 29.0 log point lower cumulative growth rate by year five; the point estimates increase in magnitude monotonically with the horizon. Relative to the unconditional mean of one-year profit growth (4.8 log points), the one-year association is quantitatively sizable, and the five-year association is several times larger than the corresponding positive association between the firm's own innovation value and profit growth (coefficient about 0.073 at year five in the same table). The pattern is consistent with persistent technological displacement: innovations that arrive in the technology space close to the firm need not reverse quickly, and slow adjustment can cumulate into multi-year profit erosion.

The coefficient on the asymmetry component is small and indistinguishable from zero at every horizon (between -0.006 and 0.000), consistent with profit growth responding to the level of displacement summarized by composite IDE rather than to the orthogonal process-product tilt. Section 4.2 shows that the asymmetry component nevertheless loads on physical and intangible capital in directions consistent with the process channel's factor-specific cost effect. The number of firm-years falls as the forward window lengthens, from 50,028 at one year to 36,963 at five. Re-estimating on the balanced panel of firms observed at every horizon leaves the coefficient path essentially unchanged—from -0.117 at one year to -0.299 at five—so the growth in magnitude with the horizon does not reflect the changing composition of the sample across columns.

4.2 Broader growth outcomes

We next estimate regressions analogous to Equation (19) for a broader set of firm performance measures: revenue, market share, employment, and capital stock (physical and intangible). The model rationalizes declines in scale and factor demand along each channel (Appendix D.5 and Appendix D.6); empirically, composite IDE inherits the predictive content common to IDE^{process} and IDE^{product} , while the asymmetry component isolates the orthogonal process-product tilt. We report estimates for both components. We define a firm's market share as the ratio of sales over the

total sales of all firms in the same 3-digit SIC industry; intangible capital follows [L. Eisfeldt, T. Kim and Papanikolaou \(2022\)](#).

Insert Tables 7 and 8 here.

Composite displacement and top-line outcomes. Table 7, columns (1)–(5), and Table 8, columns (1)–(5), report the analog of Equation (19) for revenue, market share, employment, physical capital, and intangible capital. At the five-year horizon, a one-standard-deviation increase in composite IDE is associated with declines of 30.8 log points in revenue (Panel A of Table 7), 31.8 log points in market share (Panel B of Table 7), 23.9 log points in employment (Panel A of Table 8), 44.1 log points in physical capital (Panel B of Table 8), and 36.9 log points in intangible capital (Panel C of Table 8). Revenue and market share share a numerator, so the market-share result is best read as revenue measured relative to industry peers; because the denominator varies at the industry-year level, which the fixed effects do not absorb, it is a distinct outcome rather than a rescaling of the revenue panel. The one-year coefficients (between -0.075 for employment and -0.146 for physical capital) are comparable in size to the corresponding profit-growth coefficient in Section 4.1 for revenue, market share, and intangible capital, smaller for employment, and larger for physical capital. Revenue and market share therefore move with profits in roughly the proportions implied by constant-markup pricing (Appendix D.4, Equation (46)), while intangible and especially physical capital decline by more than profits over five years—consistent with displacement lowering the marginal product of technology-specific capital and, through standard q -theoretic logic, depressing investment over multiple years.

Asymmetry component and factor demand. The asymmetry component is essentially zero for revenue (-0.008 at year five), market share (-0.003), and employment (-0.005). It is negative for physical capital, ranging from -0.003 to -0.010 at horizons of one to five years, and negative for intangible capital at every horizon, ranging from -0.005 to -0.013 at horizons of one to five years. The asymmetry-component coefficients are an order of magnitude smaller than the corresponding composite-IDE coefficients, consistent with the asymmetry component carrying roughly 9% of the within-sector-year variance of $(IDE^{\text{process}}, IDE^{\text{product}})$ —non-trivial but small relative to the

level component (Section 2.5). A positive value of the asymmetry component indicates that process-side displacement exceeds product-side displacement; the negative coefficients on capital imply that firms tilted toward process-side pressure reduce physical and intangible capital faster than the composite alone would predict—the sign pattern the process channel attaches to factors complementary to displaced technologies once investment responds (the q -theoretic margin above).

Reading the decomposition. In sum, composite IDE carries most of the predictive content of IDE for firm performance: its associations with profits, revenue, market share, employment, and capital line up with a single displacement shock that depresses scale and investment together. The asymmetry component contributes incremental predictive power for physical and especially intangible capital, in the direction the process channel implies once investment responds (the q -theoretic margin above)—process-tilted displacement reducing capital faster than the level component alone would predict. We read this as evidence consistent with the process channel operating in factor demand, while recognizing that the asymmetry component carries a modest share of within-cell variance, so the magnitude (though not the sign pattern) of the asymmetry effect should be read with some care.

4.3 Mechanisms

The preceding subsections document that composite IDE is associated with weaker firm performance across profit and non-profit outcomes, while the asymmetry component loads primarily on capital. The model points to two forward-looking objects that should comove with displacement: tightening of technology-specific labor markets (the process channel) and product obsolescence (the product channel). We project SemanticAxis risk scores from 10-K text onto composite IDE and the asymmetry component. The level of displacement summarized by composite IDE should activate both intermediate margins, while the channel tilt summarized by the asymmetry component is orthogonal to the level and need not load on either margin in particular. The two exercises therefore test whether the level comoves with both forward-looking margins, with the asymmetry component providing a benchmark for outcomes that should depend on displacement intensity rather than on process- versus product-side composition. The exact 10 sentence pairs for each axis are listed verbatim in Appendix A.5.6 and Appendix A.5.4.

4.3.1 Specialist labor market tightening

We use the SemanticAxis approach of [An, Kwak and Ahn \(2018\)](#) to measure specialist labor market tightening from 10-K filings. We instruct an LLM to extract specialist labor market risk discussions from each 10-K, embed the resulting summaries, and project them onto a tightening–easing axis constructed from 10 pairs of sentences with opposite meanings (e.g., “Hourly rates for certified technicians increased” vs. “decreased”). The first principal component of these 10 axes defines the specialist labor market tightening score $\text{splst risk}_{f,t}$ (the sentence pairs are in [Appendix A.5.6](#)).

We estimate regressions of the form:

$$\text{splst risk}_{f,t+k} = \beta_1^{L,k} \overline{\text{IDE}}_{f,t} + \beta_2^{L,k} \text{IDE}_{f,t}^{\text{tilt}} + \beta_{\text{lag}}^{L,k} \text{splst risk}_{f,t-5} + \lambda^{k\top} \mathbf{X}_{f,t} + \delta_{s \times t}^{L,k} + \gamma_n^{L,k} + \epsilon_{f,t}^{L,k}, \quad (20)$$

where we control for the firm’s own specialist labor market risk five years prior to absorb firm-level differences in reporting style.

Insert Table 9 here.

Table 9 reports a positive association between composite IDE and future specialist labor market tightening at every horizon: a one-standard-deviation increase in composite IDE is associated with about a 0.23 standard-deviation increase in one-year-ahead tightening risk in columns (1)–(3), with broadly similar magnitudes at two and three years; the coefficients rise modestly to about 0.28 across columns (4)–(6) once we additionally control for the firm’s own specialist-labor risk score five years earlier, indicating that the predictive content of composite IDE survives when we partial out the firm’s pre-existing risk-disclosure level. The coefficient on the asymmetry component is small (between -0.012 and -0.001) and indistinguishable from zero. The pattern is consistent with the process channel raising the price of technology-specific labor; empirically, it is composite IDE that covaries with the labor-market text, while the orthogonal asymmetry component is small and indistinguishable from zero on this outcome.

4.3.2 Product obsolescence risk

We apply the same SemanticAxis approach to measure product obsolescence risk from 10-K filings, using sentence pairs about high vs. low technological obsolescence (e.g., “The firm is exposed to significant technology obsolescence risk” vs. “minimal”). We focus on product-side obsolescence, which is present in about 83 percent of filings (the sentence pairs and extraction pipeline are in Appendix A.5).

We estimate analogous regressions:

$$\text{ob risk}_{f,t+k} = \beta_1^{O,k} \overline{\text{IDE}}_{f,t} + \beta_2^{O,k} \text{IDE}_{f,t}^{\text{tilt}} + \beta_{\text{lag}}^{O,k} \text{ob risk}_{f,t-5} + \lambda^{k\top} \mathbf{X}_{f,t} + \delta_{s \times t}^{O,k} + \gamma_n^{O,k} + \epsilon_{f,t}^{O,k}, \quad (21)$$

where the specification mirrors (20) but with product obsolescence risk as the dependent variable, and columns (4)–(6) control for the firm’s own obsolescence risk five years prior.

Insert Table 10 here.

Table 10 reports the analog for product obsolescence risk. Composite IDE enters positively at every horizon (coefficients between 0.11 and 0.16 in columns (1)–(3) and between 0.23 and 0.26 in columns (4)–(6), which are estimated on the smaller sample for which the five-year lagged own score is available), while the asymmetry component is small and indistinguishable from zero. The product channel of the model rationalizes obsolescence language in 10-Ks following displacement; as in the labor-market regression, it is composite IDE that covaries with the forward risk score.

Firms with high composite IDE therefore go on to disclose both tighter specialist labor markets and higher product obsolescence risk. That the asymmetry component is essentially zero on these scores is consistent with our reading that intermediate margins in the 10-K text respond to the composite *level* of displacement rather than to the orthogonal process–product tilt; it does not indicate that only one model channel is operative. Our analysis should thus be read as a joint test of the two intermediate margins at the displacement level rather than as separate identification of either channel from these projections alone.

5 Incremental value of IDE

We relate IDE to two existing displacement proxies. We use a text-based product-market similarity score, built from our 10-K embeddings following [Hoberg and Phillips \(2016\)](#) to examine the interaction between composite displacement and product-market crowding in predicting profit declines. We use the patent-citation-based technological obsolescence measure of [Ma \(2025\)](#) as a comparison on the patenting subsample, asking whether IDE retains predictive content beyond a citation-based measure of the obsolescence of the external patents a firm builds on.

5.1 IDE and Product Market Similarity

We next ask how Innovation Displacement Exposure interacts with the intensity of product-market rivalry, summarized by the Product Market Similarity ($\text{PMS}_{f,t}$) score, built from our 10-K embeddings following [Hoberg and Phillips \(2016\)](#) introduced in Section 2.4 (Equations (15) and (16)). The two measures capture distinct forces. $\text{PMS}_{f,t}$ records how crowded the focal firm’s product space is in 10-K-text similarity—that the firm sits among many close peers in product-market space, rather than that product rivalry mechanically lowers prices on its own. IDE, by contrast, records exposure to other firms’ *technological* innovations relevant to the focal firm’s process and product technologies—a forward-looking displacement pressure that need not yet be reflected in the firm’s cross-sectional product positioning.

The two forces need not act on profits in isolation. The strength with which technological displacement cumulates into profit declines should depend on how much pricing flexibility the firm retains: a firm in a crowded product space has limited room to absorb a productivity-eroding shock through markups, so a given increase in $\overline{\text{IDE}}_{f,t}$ is more likely to translate into profit losses than for an otherwise comparable firm in a thinly populated product space, which can offset some of that pressure on its existing products. Our object of interest in this subsection is therefore the *interaction* between composite displacement and product-market crowding, not a horserace between two displacement proxies; we ask whether the profit response to composite displacement steepens when the firm operates in a crowded product space, while continuing to report the IDE main effects as a check that they do not collapse once $\log \text{PMS}_{f,t}$ is held fixed.

In all regressions, we use $\log \text{PMS}_{f,t}$, standardized over the full sample and then winsorized at

the 1 percent level on each tail. We augment the baseline profit-growth regression with the level of $\log \text{PMS}_{f,t}$ and with its interactions with composite displacement and the asymmetry component:

$$\begin{aligned} \log \Pi_{f,t+k} - \log \Pi_{f,t} = & \beta_1^k \overline{\text{IDE}}_{f,t} + \beta_2^k \text{IDE}_{f,t}^{\text{tilt}} + \beta_3^k \log \text{PMS}_{f,t} \\ & + \beta_4^k \overline{\text{IDE}}_{f,t} \times \log \text{PMS}_{f,t} + \beta_5^k \text{IDE}_{f,t}^{\text{tilt}} \times \log \text{PMS}_{f,t} \\ & + \lambda^{k\top} \mathbf{X}_{f,t} + \delta_{s \times t}^k + \gamma_n^k + \epsilon_{f,t}^k. \end{aligned} \quad (22)$$

Here, the dependent variable, the control vector $\mathbf{X}_{f,t}$, the sector-by-year fixed effects $\delta_{s \times t}^k$, and the 3-digit SIC industry fixed effects γ_n^k are identical to those in Equation (19); $\overline{\text{IDE}}_{f,t}$, $\text{IDE}_{f,t}^{\text{tilt}}$, and $\log \text{PMS}_{f,t}$ all enter standardized to mean zero and unit variance over the full sample. Standard errors are clustered by firm. The moderator is interacted with the two IDE components but not with the control vector, so the interaction is identified under the assumption that the controls' associations with profit growth do not themselves vary with product-market similarity. Appendix E.3 reports a variant that interacts every control with the moderator as well; the interaction coefficients are essentially unchanged. The coefficients of interest are β_4^k and β_5^k , which test whether the profit-growth response to composite displacement and to the asymmetry component varies with how crowded the firm's product space is.

Insert Table 11 here.

Table 11 reports the estimates of Equation (22). Because $\overline{\text{IDE}}_{f,t}$, $\text{IDE}_{f,t}^{\text{tilt}}$, and $\log \text{PMS}_{f,t}$ are standardized, the coefficient on $\overline{\text{IDE}}_{f,t}$ gives the association between composite displacement and future profit growth for a firm with average Product Market Similarity. At the average level of Product Market Similarity, a one-standard-deviation increase in composite displacement is associated with a 12.2 log point lower one-year profit growth rate and a 33.2 log point lower cumulative five-year profit growth rate. The coefficient on Product Market Similarity itself is small and statistically insignificant at every horizon, indicating that crowdedness in product space does not by itself explain future profit growth once IDE, controls, and fixed effects are included.

The interaction terms show that the association between IDE and future profit growth is more negative for firms in more crowded product spaces. At the one-year horizon, moving from average

Product Market Similarity to one standard deviation above the mean makes the IDE coefficient more negative by 1.3 log points, from -0.122 to -0.135 . By year five, the same comparison makes the IDE coefficient more negative by 3.9 log points, from -0.332 to -0.371 . Equivalently, the implied five-year association of a one-standard-deviation increase in IDE is about -0.293 for firms one standard deviation below the mean of Product Market Similarity, compared with -0.371 for firms one standard deviation above the mean. Thus, product-market crowdedness amplifies the adverse association between innovation displacement and subsequent profit growth, but the amplification is modest relative to the large main association of composite IDE. The asymmetry component, $IDE_{f,t}^{\text{tilt}}$, remains economically small throughout: both its direct coefficient and its interaction with Product Market Similarity are close to zero at all horizons. This pattern reinforces the interpretation that profit growth loads primarily on the overall level of displacement exposure, rather than on whether the exposure is tilted toward process or product technologies.

5.2 IDE and patent-citation-based technological obsolescence

[Ma \(2025\)](#) measures technological obsolescence using the decline in external citations received by the externally-sourced patents a firm builds on—those it cites, excluding its own. This measure differs from IDE in two ways. First, it is restricted to patenting firms (roughly 30 percent of our sample). Second, it captures the obsolescence of the external patents the firm cites and builds on rather than exposure to external innovations affecting the firm’s full technology base. We use the five-year look-back variant from [Ma \(2025\)](#), which is his main measure and matches our IDE look-back window. We merge by year and `gvkey`, then standardize and winsorize the score at the 1 percent level on each tail.

Insert Table 12 here.

Table 12 presents two panels. Panel A estimates our baseline specification on the subsample of patenting firms. In this subsample, composite IDE remains strongly negatively associated with profit growth (coefficients of about -0.079 at year one and -0.275 at year five), while the asymmetry component is small and indistinguishable from zero—again consistent with the level of displacement driving top-line profit dynamics on this subsample. Panel B adds the citation-based

obsolescence measure as a control. The composite-IDE coefficients are essentially unchanged, and the citation-based measure is negative and statistically significant only at the one-year horizon, and indistinguishable from zero at every longer horizon. The table therefore indicates that IDE carries displacement information not contained in the citation-based measure alone.

6 IDE and industry-level outcomes

The preceding sections establish that higher composite IDE is associated with weaker firm-level performance. We now turn to the industry: does the same exposure that depresses incumbent profits raise the average level of industry productivity, or does it mostly redistribute activity across firms within an industry? As discussed in the introduction, the two leading views of creative destruction—whether creative destruction raises overall growth or it primarily leads to reallocation—are not mutually exclusive, but they predict different industry-level patterns. Higher composite displacement exposure should raise average industry productivity if the growth view dominates; it should raise dispersion with little or no impact on average growth, if the reallocation view dominates.

Section 6.1 asks whether composite displacement raises the average level of industry productivity. Section 6.2 asks whether composite displacement reorganizes activity within an industry, on two reallocation margins—the cross-firm dispersion of firm productivity growth.

Industry-level aggregation. We first aggregate our product and process IDE measures at the 3-digit SIC level. We compute asset-weighted industry means, restricting to industry-year cells with at least five IDE-valid firms. We re-extract the first two principal components on the industry-year panel. Since the industry-level variables are highly skewed, we transform them into their empirical percentile ranks in $[0, 1]$; thus, a one-unit change in $\overline{\text{IDE}}_{i,t}$ is the move from the 0th to the 100th percentile of the industry-year distribution. Appendix A.3 contains further details.

6.1 Average industry productivity

We first ask whether creative destruction raises the average level of industry productivity. We focus on two complementary sources of data. First, we focus on the sample of Compustat firms:

for each 3-digit SIC industry, we compute the growth in aggregate revenue, or gross profit, per worker—defined as the ratio of the total sales (or gross profit) of all firms in that industry divided by their total employment. An advantage of this sample is that it allows us to also measure dispersion in firm growth rates, as we discuss below. The disadvantage is that it is restricted to publicly traded firms. Hence, as a second source of data we also consider the growth in labor productivity at the 4-digit NAICS level from the Bureau of Economic Analysis (BEA). The BEA data cover the full universe of producers—public and private—and remove any concern that the Compustat aggregate is mechanically driven by the same firms that build the IDE measure. The disadvantage of the BEA data is that they do not contain sufficient information for us to compute measures of reallocation—dispersion in firm growth rates.

We estimate, for each horizon $k \in \{1, \dots, 5\}$,

$$\log Y_{i,t+k} - \log Y_{i,t} = \beta_{1,k} \overline{\text{IDE}}_{i,t} + \beta_{2,k} \text{IDE}_{i,t}^{\text{tilt}} + \gamma_k \log Y_{i,t} + \alpha_t + \varepsilon_{i,t,k}, \quad (23)$$

where $Y_{i,t}$ is measured productivity in industry i in year t , and $\overline{\text{IDE}}_{i,t}$ and $\text{IDE}_{i,t}^{\text{tilt}}$ are the innovation displacement exposure measures transformed into industry-year percentile ranks as discussed above. We include year fixed effects α_t , where year denotes the Compustat fiscal year throughout the paper, and standard errors are clustered at the industry level.

Insert Tables 13 and 14 here.

Table 13 focuses on the Compustat sample while Table 14 focuses on the BEA sample. Across both samples, both productivity definitions, and all horizons, we find no evidence that composite displacement raises subsequent growth in industry productivity. The asymmetry component is also weak in absolute magnitude and statistically insignificant at every horizon in both series. In sum, we find no evidence that the degree of creative destruction in an industry raises average growth in industry productivity.

6.2 Dispersion in firm growth

We next ask whether creative destruction reorganizes activity within an industry rather than lifting the average. Here we are restricted to the Compustat sample, because estimating the within-industry distribution of firm-level growth requires firm-level observations.

We estimate the following specification for each horizon $k \in \{1, \dots, 5\}$:

$$\text{Spread}_{i,t}(\Delta_k \log y_f) = \beta_{1,k} \overline{\text{IDE}}_{i,t} + \beta_{2,k} \text{IDE}_{i,t}^{\text{tilt}} + \gamma_k \text{Spread}_{i,t}(\log y_f) + \alpha_t + \varepsilon_{i,t,k}, \quad (24)$$

where $\text{Spread}_{i,t}(\Delta_k \log y_f)$ is the within-industry P90–P10 spread of firm-level k -year log growth in measured productivity in industry i at date t . We consider two measures of firm-level productivity, y_f : revenue per employee, $\text{sale}_{f,t}/\text{emp}_{f,t}$, and gross profit per employee, $(\text{sale}_{f,t} - \text{cogs}_{f,t})/\text{emp}_{f,t}$. The control $\text{Spread}_{i,t}(\log y_f)$ is the contemporaneous within-industry P90–P10 spread of the corresponding log level. $\overline{\text{IDE}}_{i,t}$ and $\text{IDE}_{i,t}^{\text{tilt}}$ are the industry-level innovation displacement exposure measures transformed into industry-year percentile ranks, as discussed above. As before, we include year fixed effects α_t and cluster standard errors at the industry level.

Insert Table 15 here.

Panels A and B of Table 15 report the estimated coefficients; Panel A focuses on dispersion in the growth in revenue per employee, while Panel B focuses on the dispersion in the growth in gross profit per employee. Examining the table, we see that the composite displacement measure is both statistically and economically significantly related to subsequent dispersion in within-industry firm outcomes. In terms of magnitudes, a 0-to-100-percentile move in $\overline{\text{IDE}}_{i,t}$ widens the P90–P10 spread of one-year log revenue-per-employee growth by 0.57, and by between 0.65 and 0.79 at horizons of two to five years. The results in Panel B are qualitatively similar though somewhat weaker in magnitude. By contrast, the asymmetry component plays no role in dispersion at any horizon.

In sum, we find no evidence that creative destruction raises the level of industry productivity in our sample, but it does lead to increased dispersion in firm outcomes.

7 Robustness

We assess the robustness of our main profit-growth results in Table 6 along five dimensions. First, we include placebo IDE measures constructed by randomizing the fiscal year ordering of each firm’s annual technology summaries within the five-year window, which breaks temporal alignment while preserving cross-sectional structure in technology space. Second, we attach forward-dated IDE values to current firm-years, a measure that could carry predictive power only if the summaries encoded information post-dating the filing. Third, we drop innovation-value weights and re-form composite displacement from raw cosine sums. Fourth, we residualize composite displacement and the asymmetry component on an expanded control set, and lastly, we decompose exposure into moderate- and high-similarity innovator tails relative to the global distribution of pairwise cosine similarities. Additional exercises (IDE interacted with firm characteristics) remain in Appendix C.

7.1 Placebo tests

Both placebos in this subsection target the same underlying concern—that the headline result reflects slow-moving cross-sectional structure in the innovation landscape rather than the year alignment between a firm’s evolving technology disclosures and other firms’ contemporaneous innovations. We break the alignment in two complementary directions: first by randomly permuting the year ordering of each firm’s technology-stack inputs while holding the innovation side fixed, and second by attaching forward-dated IDE values to current firm-years. The first asks whether the correctly timed measure survives its shuffled analog, which under the null that timing is irrelevant would absorb its predictive content; the second asks whether a measure built from filings dated after t predicts profit growth realized from t , which it could do only through leakage. The data support the correctly timed measure in both cases.

7.1.1 Placebo 1: randomizing the timing of technology stacks

10-K technology disclosures evolve slowly, so within-firm temporal variation in the technology stack might carry little independent signal beyond the slow-moving cross-section. We construct shuffled-stack placebo versions of process and product IDE in which, within each firm, the year ordering of the five-year technology-stack inputs is randomly permuted while holding the innovation side

fixed, and re-estimate the baseline profit-growth regression including both the correctly timed composite displacement and asymmetry components and the placebo analogs:

$$\begin{aligned} \log \Pi_{f,t+k} - \log \Pi_{f,t} = & \beta_1^k \overline{\text{IDE}}_{f,t} + \beta_2^k \text{IDE}_{f,t}^{\text{tilt}} + \beta_3^k \overline{\text{IDE}}_{f,t}^{\text{shuffle}} + \beta_4^k \text{IDE}_{f,t}^{\text{tiltshuffle}} \\ & + \lambda^{k\top} \mathbf{X}_{f,t} + \delta_{s \times t}^k + \gamma_n^k + \epsilon_{f,t}^k, \end{aligned} \quad (25)$$

where $\overline{\text{IDE}}_{f,t}^{\text{shuffle}}$ and $\text{IDE}_{f,t}^{\text{tiltshuffle}}$ denote the composite displacement and asymmetry components formed by applying the same principal-component convention as in Section 2.5 to channel-specific IDE measures computed using randomly shuffled technology-stack timing within firm f , with process and product channels shuffled separately before forming the PCs. Under the null that timing is irrelevant, the placebo analogs should absorb the variation that drives our results.

Insert Table 16 here.

Table 16 sets the two specifications side by side: Panel A is the baseline and Panel B adds the placebo analogs as controls. The correctly timed coefficients are close to unchanged between them, moving from -0.106 to -0.118 at one year and from -0.290 to -0.276 at five, so a one-standard-deviation increase in composite displacement is associated with about an 11.8 log point decline in one-year profit growth and about a 27.6 log point decline by year five once the placebos are held fixed. The asymmetry component is small throughout in both panels, and statistically insignificant at every horizon other than the first, where it is marginally significant. This is a demanding comparison rather than a formality: because 10-K disclosures evolve slowly, permuting a firm's own stack ordering barely changes the resulting measure — the shuffled and correctly timed composites correlate at 0.98 — so the placebo is a near-collinear control. That the correctly timed coefficients survive it indicates the predictive content is not an artifact of arbitrary within-firm ordering of the annual technology summaries.

7.1.2 Placebo 2: an IDE with lookahead bias

A second concern is that the LLM-derived summaries could encode information that post-dates the filing, so that similarity to innovations lines up with subsequent outcomes for reasons unrelated to

displacement pressure at date t . The direct test is to give the regression a measure that could only help if such leakage were present. We attach to firm-year (f, t) the IDE measured at $(f, t + h)$ and re-estimate the baseline profit-growth regression for each h :

$$\log \Pi_{f,t+k} - \log \Pi_{f,t} = \beta_1^k \overline{\text{IDE}}_{f,t+h} + \beta_2^k \text{IDE}_{f,t+h}^{\text{tilt}} + \lambda^{k\top} \mathbf{X}_{f,t} + \delta_{s \times t}^k + \gamma_n^k + \epsilon_{f,t}^k, \quad (26)$$

where the forward-dated regressors $\overline{\text{IDE}}_{f,t+h}$ and $\text{IDE}_{f,t+h}^{\text{tilt}}$ replace their contemporaneous counterparts in Equation (19), while the control vector $\mathbf{X}_{f,t}$, the outcome window and the fixed effects are unchanged and remain dated at t . The sample is the baseline estimation sample, so the only difference from Equation (19) is which vintage of the measure enters. Because IDE at date s is built from a five-year look-back over filings $[s - 4, s]$, the measure at $t + h$ draws on filings dated after t once $h \geq 5$; we therefore report $h = 5$, the first vintage with no overlap, and $h = 7$. Under the null that the measure reflects contemporaneous exposure, a forward-dated IDE should carry no predictive power for profit growth realized between t and $t + k$; if the summaries leaked future information, it should predict at least as well as the contemporaneous measure.

Insert Table 17 here.

Table 17 shows that the forward-dated measure carries little predictive content. At $h = 5$ the composite coefficient is -0.037 at the one-year horizon, marginally significant, against -0.106 for the contemporaneous measure, and it is statistically indistinguishable from zero at every longer horizon. At $h = 7$ it is indistinguishable from zero at every horizon, including the first. The asymmetry component is indistinguishable from zero throughout. The point estimates run at roughly a third of the contemporaneous ones and rise with the outcome horizon, which is what the persistence of the measure implies: IDE at $t + 5$ is correlated with IDE at t , so a forward-dated measure retains some of the contemporaneous signal. What it does not do is predict as well as the contemporaneous measure, which is what look-ahead in the summaries would produce.

7.2 IDE without innovation-value weights

Our baseline IDE weights cosine similarity by other firms' innovation values (scaled by the innovating firms' assets), following the model's emphasis on economically large external innovations. To assess how much of the predictive content comes from the similarity machinery alone, we drop the weights and sum raw cosines between each focal firm's process and product technology embeddings and every other firm's innovation embedding:

$$\begin{aligned} IDE_{f,t}^{\text{process,raw}} &= \sum_{g \neq f}^{n_{t,\text{Innov}}} \cos(\text{Innov}_{g,t}, \text{Tech}_{f,t}^{\text{process}}), \\ IDE_{f,t}^{\text{product,raw}} &= \sum_{g \neq f}^{n_{t,\text{Innov}}} \cos(\text{Innov}_{g,t}, \text{Tech}_{f,t}^{\text{product}}). \end{aligned} \quad (27)$$

We then form composite displacement and the asymmetry component from the standardized pair $(IDE_{f,t}^{\text{process,raw}}, IDE_{f,t}^{\text{product,raw}})$ exactly as in Section 2.5 and re-estimate Equation (19).

Insert Table 18 here.

Table 18 reports the estimates. Dropping the weights attenuates the composite coefficient, but composite displacement remains negative at every horizon and the asymmetry component is small throughout. The weights change the scale of the estimated response rather than the fit: the incremental R^2 of the IDE terms over a specification with the controls and fixed effects alone is 0.005 at five years with the weights and 0.006 without them, and the unweighted measure is estimated no less precisely. At the five-year horizon, the point estimate is roughly 40 percent of the magnitude in Table 6, so the qualitative pattern survives even when every innovator enters the sum with equal weight.

7.3 Residual IDE with expanded controls

Section 2.4 shows that standard firm characteristics explain only about 12 percent of the conditional variation in process and product IDE. We next ask whether composite displacement retains incremental predictive content for profit growth when we partial out *all* of that covariation aggressively. We expand the baseline control vector in Equation (19) to include firm age, Product Market Sim-

ilarity, and variable cost (which reduces the sample by over 30 percent and introduces potential selection concerns, so we treat this as a robustness check rather than our preferred specification). We project $\overline{IDE}$ and IDE^{tilt} on the full expanded control set and use the residuals as regressors, attributing any overlap in profit-related information between IDE and the controls entirely to the controls.

Insert Table 19 here.

Table 19 reports the residualized specification. Residual composite displacement remains negative at every horizon and broadly matches the baseline magnitudes in Table 6; the asymmetry component is small throughout. The comparison holds despite the smaller sample—29,418 firm-years at $t + 1$ versus 50,028 in the baseline table—which indicates that composite displacement carries information about subsequent profit growth not spanned by the expanded control set alone.

7.4 IDE from highly similar innovators

Finally, we ask whether the headline result is concentrated in the upper tail of the firm-pair similarity distribution—innovators whose embeddings lie closest to the focal firm—or whether the broader sum in Equation (14) contains incremental information. We split innovator pairs into “high” exposure (cosine similarity at least two standard deviations above the mean similarity across all firm-pairs and years) and “moderate” exposure (at least one standard deviation above the mean), form composite displacement and the asymmetry component within each bucket using the same PCA convention as in Section 2.5, and re-estimate the baseline specification in three layers reported in Table 20.

Insert Table 20 here.

Panel A includes only the high-similarity components. Composite displacement in the high tail is negative and statistically significant at every horizon, from -0.008 at one year to -0.047 at five. Panel B adds the moderate-similarity components while controlling for the high-similarity components; the moderate-similarity composite is negative and significant at every horizon and larger in magnitude than the high-tail composite it conditions on, so displacement risk is not

confined to the extreme tail. Panel C includes the baseline composite displacement and asymmetry component while controlling for the moderate-similarity components; the baseline composite remains negative and significant throughout, from -0.155 to -0.335 . Composite displacement is therefore negatively associated with subsequent profit growth at each level of the similarity distribution at which it is measured, and at each level the association survives conditioning on the others.

What distinguishes the buckets from one another is economically modest by comparison. The incremental components—the high-similarity composite in Panel B and the moderate-similarity composite in Panel C—are much smaller than the composite they are entered alongside—typically around a tenth of its magnitude, and at most about half—are only intermittently significant, and enter with the opposite sign once the broader measure is included. That is what nested buckets imply: the broader measure already summarizes the shared variation, so the narrower one retains little independent content. The asymmetry components are indistinguishable from zero at nearly every horizon in all three panels.

8 Conclusion

We construct firm-level Innovation Displacement Exposure (IDE) from the similarity between other firms' patent-based innovation embeddings and the focal firm's 10-K technology descriptions, weighted by aggregate patent value, and summarize the model's process and product analogs by a composite IDE measure and an orthogonal asymmetry component. Relative to industry-peer innovation exposure ([Kogan et al., 2017](#)), IDE matches rivals' innovations to the focal firm's own product and technology stacks and separates product-market from process-side displacement. In a panel of U.S. public firms with 10-K coverage from 1997 to 2019, a one-standard-deviation increase in composite IDE forecasts large declines in profit growth, revenue, market share, employment, and physical and intangible capital, with magnitudes that increase with the horizon. Forward-looking 10-K text measures of specialist labor-market tightening and product obsolescence risk both rise with composite IDE, consistent with displacement propagating through the process and product margins in parallel at the level summarized by the composite. The asymmetry component is indistinguishable from zero for profit growth and for those mechanism scores but loads negatively

on physical and especially intangible capital in directions consistent with the process channel. At the industry level, percentile-ranked composite displacement comoves with wider within-industry dispersion in firm growth but not with higher average industry productivity in Compustat or BEA series—in our sample, creative destruction at the industry level manifests as reallocation, not as a level shift.

Our work opens up several avenues for future research. First, IDE can be embedded in asset-pricing tests to ask whether displacement risk is priced in the cross-section of equity returns. Second, extending the measurement pipeline to private firms and to non-U.S. markets with 10-K-equivalent disclosures would probe the external validity of the displacement gradients we estimate among listed U.S. incumbents. Last, a structural model that disciplines dynamic investment and factor-demand responses to external innovation shocks could translate our reduced-form magnitudes into welfare and policy counterfactuals.

References

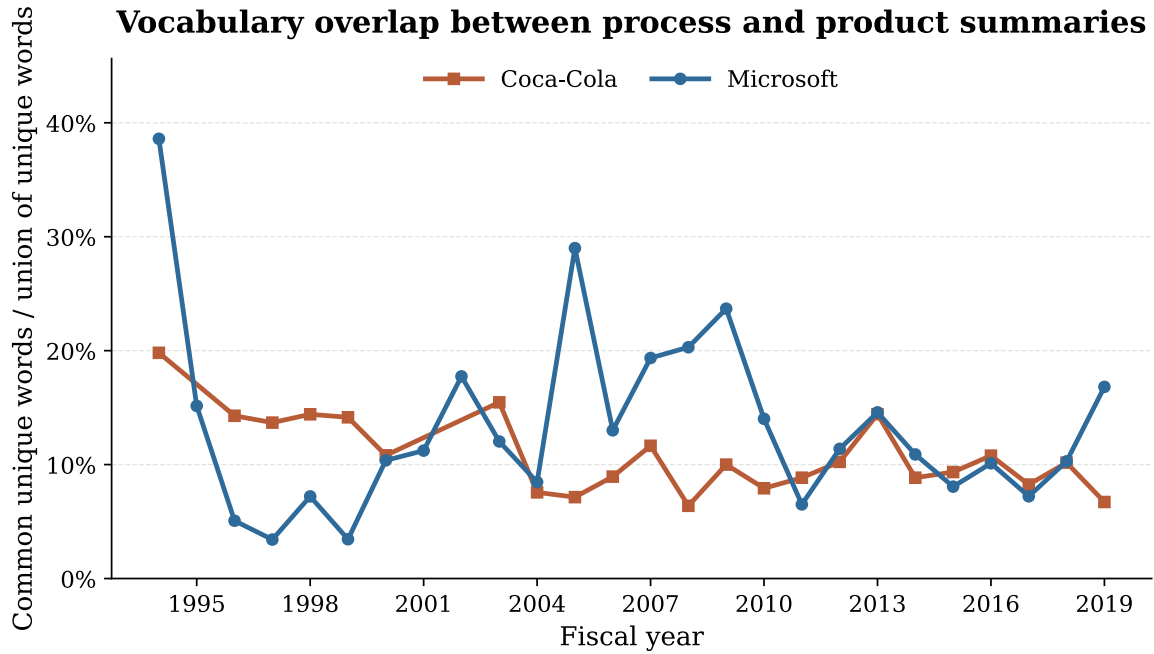
- Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr.** 2018. “Innovation, reallocation, and growth.” *American Economic Review*, 108(11): 3450–3491.
- Aghion, Philippe, and Peter Howitt.** 1992. “Growth Through Creative Destruction.” *Econometrica*, 60(2): 323–51.
- Akcigit, Ufuk, and Sina T Ates.** 2019. “What Happened to U.S. Business Dynamism?” National Bureau of Economic Research NBER Working Paper 25756.
- Alexopoulos, Michelle.** 2011. “Read all about it!! What happens following a technology shock?” *American Economic Review*, 101(4): 1144–1179.
- An, Jisun, Haewoon Kwak, and Yong-Yeol Ahn.** 2018. “Semaxis: A lightweight framework to characterize domain-specific word semantics beyond sentiment.” *arXiv preprint arXiv:1806.05521*.
- Bilir, L Kamran.** 2014. “Patent laws, product life-cycle lengths, and multinational activity.” *American Economic Review*, 104(7): 1979–2013.

- Bloom, Nicholas, Mark Schankerman, and John Van Reenen.** 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica*, 81(4): 1347–1393.
- Breitung, Christian, and Sebastian Müller.** 2025. "Global business networks." *Journal of Financial Economics*, 166: 104007.
- Caskurlu, Tolga, Gerard Hoberg, and Gordon M Phillips.** 2024. "New Technology Sectoral Disruptions." University of Amsterdam, University of Southern California and Tuck School of Business at Dartmouth Working Paper.
- Cohen, Lauren, Christopher Malloy, and Quoc Nguyen.** 2020. "Lazy Prices." *The Journal of Finance*, 75(3): 1371–1415.
- Davis, Steven J, and John Haltiwanger.** 1992. "Gross job creation, gross job destruction, and employment reallocation." *The Quarterly Journal of Economics*, 107(3): 819–863.
- Dyer, Travis, Mark Lang, and Lorien Stice-Lawrence.** 2017. "The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation." *Journal of Accounting and Economics*, 64(2-3): 221–245.
- Fedyk, Anastassia, Ali Kakhbod, Peiyao Li, and Ulrike Malmendier.** 2024. "AI and Perception Biases in Investments: An Experimental Study." *Available at SSRN* 4787249.
- Fernández-Villaverde, Jesús, Yang Yu, and Francesco Zanetti.** 2025. "Defensive Hiring and Creative Destruction." National Bureau of Economic Research NBER Working Paper 33588.
- Garcia-Macia, Daniel, Chang-Tai Hsieh, and Peter J. Klenow.** 2019. "How Destructive Is Innovation?" *Econometrica*, 87(5): 1507–1541.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy.** 2019. "Text as data." *Journal of Economic Literature*, 57(3): 535–74.
- Griliches, Zvi, and Haim Regev.** 1995. "Firm productivity in Israeli industry 1979–1988." *Journal of Econometrics*, 65(1): 175–203.
- Haltiwanger, John.** 2012. "Job Creation and Ind Dynamics in the United States." In *Innovation Policy and the Economy*, Volume 12. NBER Chapters, 17–38. National Bureau of Economic Research, Inc.

- Hansen, Stephen, Michael McMahon, and Andrea Prat.** 2018. "Transparency and deliberation within the FOMC: a computational linguistics approach." *Quarterly Journal of Economics*, 133(2): 801–870.
- Hoberg, Gerard, and Gordon Phillips.** 2016. "Text-based network industries and endogenous product differentiation." *Journal of Political Economy*, 124(5): 1423–1465.
- Jaffe, Adam B.** 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review*, 76(5): 984–1001.
- Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy.** 2021. "Measuring technological innovation over the long run." *American Economic Review: Insights*, 3(3): 303–20.
- Klette, Tor Jakob, and Samuel Kortum.** 2004. "Innovating firms and aggregate innovation." *Journal of Political Economy*, 112(5): 986–1018.
- Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman.** 2017. "Technological innovation, resource allocation, and growth." *Quarterly Journal of Economics*, 132(2): 665–712.
- Kogan, Leonid, Dimitris Papanikolaou, and Noah Stoffman.** 2020. "Left behind: Creative destruction, inequality, and the stock market." *Journal of Political Economy*, 128(3): 855–906.
- Kortum, Samuel.** 1993. "Equilibrium r&d and the patent–r&d ratio: Us evidence." *The American Economic Review*, 83(2): 450–457.
- Kortum, Samuel, and Josh Lerner.** 1998. "Stronger protection or technological revolution: what is behind the recent surge in patenting?" Vol. 48, 247–304, Elsevier.
- L. Eisfeldt, Andrea, Edward T. Kim, and Dimitris Papanikolaou.** 2022. "Intangible value." *Critical Finance Review*, 11(2): 299–332.
- Li, Feng.** 2008. "Annual report readability, current earnings, and earnings persistence." *Journal of Accounting and Economics*, 45(2-3): 221–247.
- Li, Feng.** 2010. "The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach." *Journal of Accounting Research*, 48(5): 1049–1102.

- Loughran, Tim, and Bill McDonald.** 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance*, 66(1): 35–65.
- Loughran, Tim, and Bill McDonald.** 2014. "Measuring Readability in Financial Disclosures." *The Journal of Finance*, 69(4): 1643–1671.
- Loughran, Tim, and Bill McDonald.** 2016. "Textual Analysis in Accounting and Finance: A Survey." *Journal of Accounting Research*, 54(4): 1187–1230.
- Ma, Song.** 2025. "Technological obsolescence." *The Review of Financial Studies*, Forthcoming.
- Romer, Paul M.** 1990. "Endogenous technological change." *Journal of political Economy*, 98(5, Part 2): S71–S102.
- Schumpeter, Joseph A.** 1942. *Capitalism, socialism and democracy*. routledge.
- Shea, John.** 1998. "What do technology shocks do?" *NBER macroeconomics annual*, 13: 275–310.
- You, Haifeng, and Xiao-jun Zhang.** 2009. "Financial Reporting Complexity and Investor Underreaction to 10-K Information." *Review of Accounting Studies*, 14(4): 559–586.

Figures and Tables



Note: The plotted measure is the Jaccard overlap between input/process and output/product summary vocabularies.

Figure 1: Word overlap between process and product technology summaries over time. For each firm-year, we plot the fraction of lemmatized words that appear in both summaries.

Innovation Displacement Exposure by industry

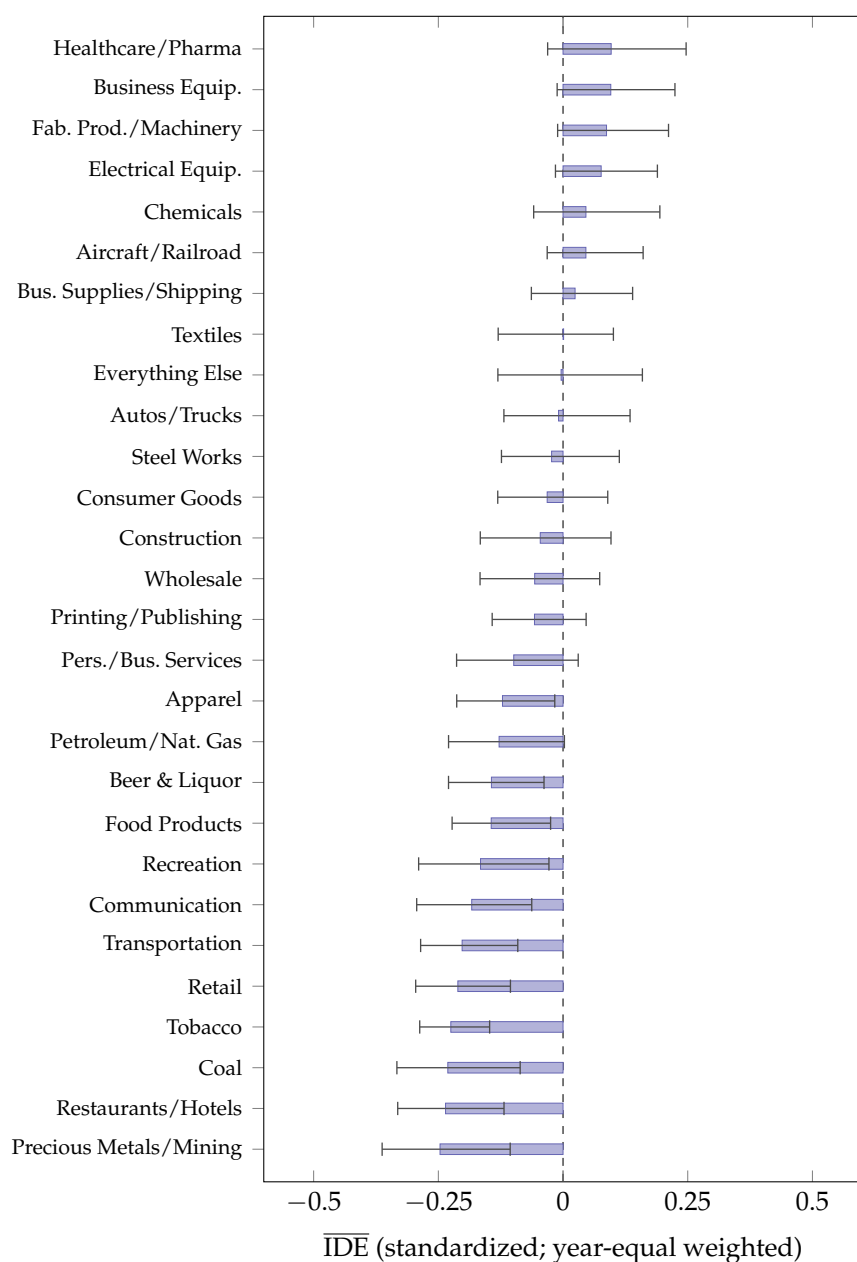


Figure 2: Innovation Displacement Exposure by industry (extended sample, 1997–2019). $\overline{\text{IDE}}$ is the standardized first principal component (PC₁) of process and product IDE. To neutralize sample-composition tilt from industries with heavy entry/exit, statistics are computed with year-equal weighting: for each industry-year we compute the cross-firm mean and the 25th/75th percentiles, then average those statistics across years. Bars show the year-equal-weighted mean; horizontal lines show the year-equal-weighted within-industry interquartile range. Industries are sorted in descending order by mean (highest at top).

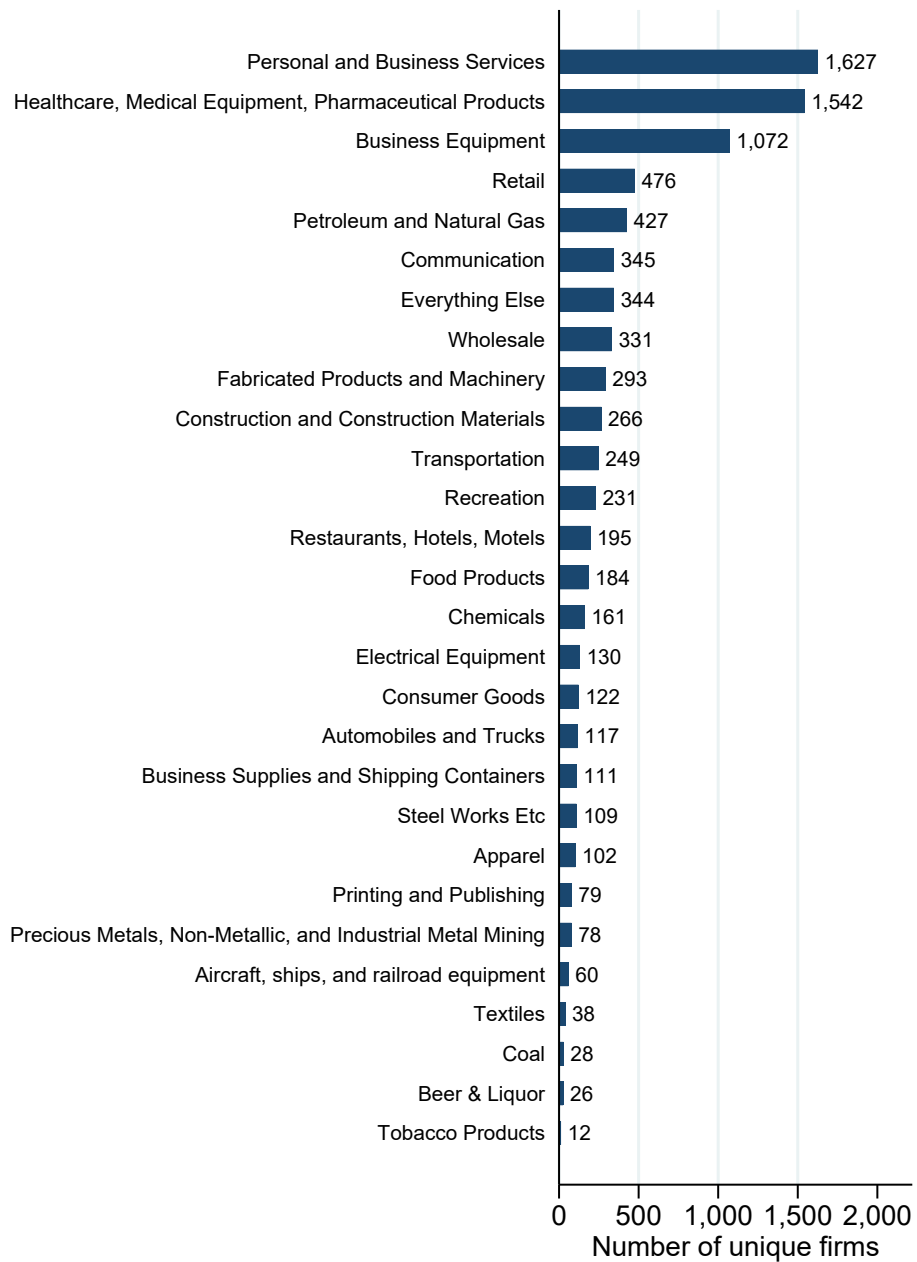


Figure 3: Number of unique firms in each sector in our sample (1997–2019).

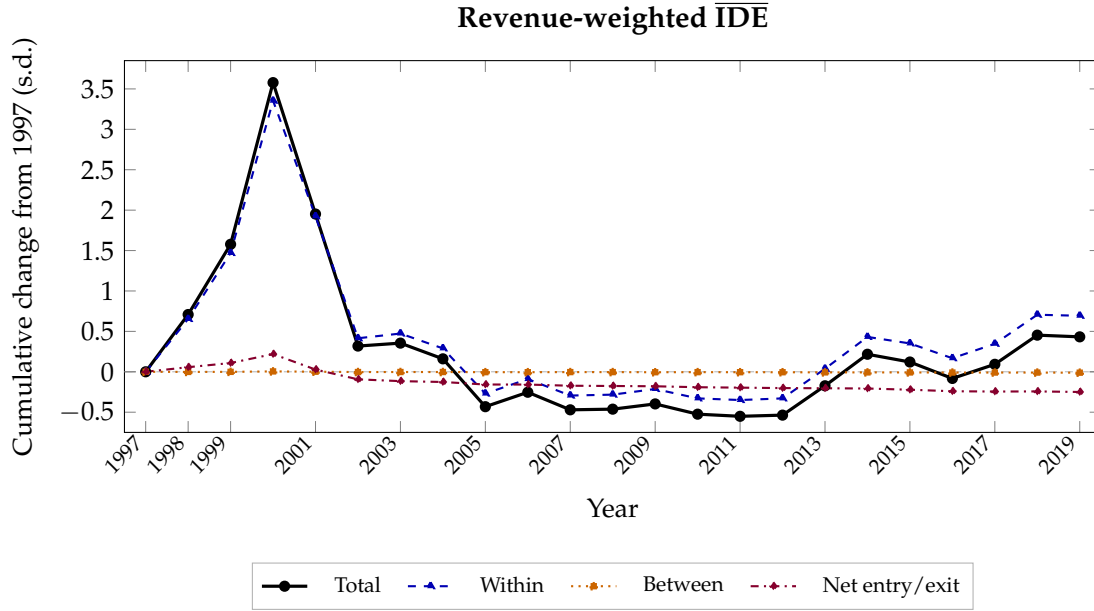


Figure 4: Decomposition of cumulative change in revenue-weighted mean $\overline{\text{IDE}}$ (1997–2019). $\overline{\text{IDE}}$ is the standardized first principal component (PC_1) of process and product IDE. The figure decomposes the cumulative change from 1997 (= 0) in the revenue-weighted cross-sectional mean of standardized $\overline{\text{IDE}}$ into three components using the [Griliches and Regev \(1995\)](#) identity. *Within*: change in continuing firms’ IDE, weighted by average revenue share. *Between*: revenue reallocation toward or away from high-IDE continuing firms (weight changes, IDE held fixed). *Net entry/exit*: combined contribution of entering and exiting firms. The three components sum exactly to the total at each point.

Process-only (top 10)	Common (top 10)	Product-only (top 10)
product (26)	microsoft (48)	firm (12)
research (26)	management (40)	use (12)
internal (25)	application (38)	support (11)
service (19)	provide (38)	communication (9)
internally (17)	operate (34)	live (9)
production (16)	software (34)	dynamic (8)
facility (15)	tool (34)	integration (8)
new (14)	window (34)	consumer (7)
manufacture (13)	productivity (30)	email (6)
solely (13)	customer (29)	excel (6)

TABLE 1: Most frequent process-only, common, and product-only lemmatized tokens (English stopwords plus corpus-wide $>80\%$ high-frequency words removed) from the technology summaries of Microsoft’s 10-Ks. Counts in parentheses report the number of annual summaries (pooling process and product) in which the token appears.

Process-only (top 10)	Common (top 10)	Product-only (top 10)
production (23)	beverage (42)	firm (19)
product (21)	distribution (40)	diet (13)
atlantic (17)	company (39)	energy (13)
canners (17)	customer (39)	dasani (10)
process (17)	sell (38)	use (10)
south (17)	summary (34)	classic (9)
internally (16)	drink (33)	yummies (9)
material (16)	market (32)	monster (8)
raw (15)	coca-cola (31)	tum-e (8)
control (13)	distribute (31)	pepper (6)

TABLE 2: Most frequent process-only, common, and product-only lemmatized tokens (English stopwords plus corpus-wide $>80\%$ high-frequency words removed) from the technology summaries of Coca-Cola’s 10-Ks. Counts in parentheses report the number of annual summaries (pooling process and product) in which the token appears.

	Raw correlation		Multivariate regression (FE-residualized)	
	Process IDE	Product IDE	Process IDE	Product IDE
Product-weighted sales growth	-0.633	-0.629	-0.147*** (0.014)	-0.180*** (0.016)
Firm-level innovation value	0.113	0.114	0.003 (0.002)	0.002 (0.003)
Log capital stock	-0.170	-0.177	0.008*** (0.002)	0.005* (0.003)
Log employment size	-0.097	-0.102	-0.005* (0.003)	-0.004 (0.003)
Variable cost	-0.002	0.002	0.007 (0.005)	0.011** (0.005)
Log current profit	-0.124	-0.130	0.001 (0.003)	0.005 (0.003)
Profitability	-0.026	-0.029	0.028*** (0.008)	0.021** (0.010)
Product Market Similarity	-0.602	-0.594	0.134*** (0.009)	0.185*** (0.010)
Age	-0.089	-0.090	0.008*** (0.000)	0.008*** (0.000)
Observations			32,271	32,271
R ²			0.127	0.122

TABLE 3: Firm characteristics associated with Innovation Displacement Exposure. The first two columns report raw correlations between each firm characteristic and standardized process and product IDE. The last two columns report coefficients from a multivariate regression of standardized IDE on all characteristics where every variable is first residualized against sector-by-year and industry fixed effects; the regression itself contains no fixed effects. Standard errors (in parentheses) are clustered by firm. The sample is 1997–2019; observations with missing values in any variable are dropped. Firm age since IPO is the binding constraint: it is available for 42,644 of the 70,236 IDE-valid firm-years, and the remaining characteristics would leave about 54,000 observations without it. One asterisk denotes statistical significance at the 10 percent level, two at 5 percent, and three at 1 percent. All firm-level variables are winsorized at the 1/99% level in each year.

TABLE 4: Summary statistics for composite IDE and the asymmetry component, standardized over the full firm-year sample (1997–2019). Unlike the regression tables, whose inputs are winsorized within each year, the inputs to this table are winsorized at 1/99% on the pooled panel. $\overline{\text{IDE}}$ (composite displacement) and IDE^{tilt} (asymmetry) are the standardized first and second principal components of process and product IDE. Skewness is the sample third-moment skewness.

	N	Mean	SD	Min	p5	p25	Median	p75	p95	p99	Max	Skew.
Composite IDE ($\overline{\text{IDE}}$)	70,236	0.00	1.00	−1.15	−0.95	−0.67	−0.24	0.20	2.22	3.73	3.80	1.92
Asymmetry (IDE^{tilt})	70,236	0.00	1.00	−7.20	−1.55	−0.57	−0.01	0.55	1.58	2.73	11.08	0.15

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Forward horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
<i>Panel A: process IDE alone</i>					
IDE ^{process}	−0.101*** (0.010)	−0.163*** (0.017)	−0.195*** (0.023)	−0.237*** (0.028)	−0.272*** (0.034)
<i>Panel B: product IDE alone</i>					
IDE ^{product}	−0.087*** (0.010)	−0.146*** (0.016)	−0.183*** (0.022)	−0.206*** (0.027)	−0.242*** (0.032)
<i>Panel C: both channels jointly</i>					
IDE ^{process}	−0.073*** (0.015)	−0.104*** (0.025)	−0.106*** (0.035)	−0.166*** (0.043)	−0.178*** (0.051)
IDE ^{product}	−0.033** (0.015)	−0.070*** (0.024)	−0.106*** (0.033)	−0.085** (0.040)	−0.113** (0.048)
Year × Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓
Observations	50,028	45,970	42,522	39,567	36,963

TABLE 5: The channel measures entered directly. Panels A and B enter process IDE and product IDE one at a time; Panel C enters both together. Both channels are standardized over the IDE-valid sample, so coefficients are per standard deviation. The two channels are highly collinear — their raw correlation is 0.99, and 0.81 within sector-by-year and industry cells, which is the variation Panel C identifies from — so the joint standard errors are roughly 50 percent larger than in Panels A and B. This is why the paper summarizes the pair by principal components rather than entering the channels separately. The specification is the baseline profit-growth regression with the composite replaced by the channel measures. The baseline controls are current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); log of current profit; current capital stock (Compustat ppeg, in logs); current employment (Compustat emp, in logs); product-weighted sales growth—the product-similarity-weighted average of contemporaneous sales growth across all other firms, where the weights are the pairwise product similarities from [Hoberg and Phillips \(2016\)](#); and aggregate firm-level innovation value, deflated by total assets. Fixed effects are sector-by-year and 3-digit SIC industry. Standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Horizon (years)				
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}} (\text{PC}_1)$	-0.106*** (0.011)	-0.173*** (0.018)	-0.211*** (0.024)	-0.250*** (0.029)	-0.290*** (0.035)
$\text{IDE}^{\text{tilt}} (\text{PC}_2)$	-0.003 (0.002)	-0.002 (0.003)	-0.000 (0.004)	-0.006 (0.005)	-0.005 (0.006)
Firm innovation value	0.027*** (0.002)	0.043*** (0.004)	0.054*** (0.006)	0.063*** (0.008)	0.073*** (0.009)
Observations	50,028	45,970	42,522	39,567	36,963
R ²	0.134	0.151	0.146	0.146	0.152
Fixed Effects:					
Year $\times$ Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓

TABLE 6: IDE and firm profit. This table shows the association between our IDE measure and profit growth over the next five years (one to five years ahead). $\overline{\text{IDE}} (\text{PC}_1)$ is the composite displacement and $\text{IDE}^{\text{tilt}} (\text{PC}_2)$ is the asymmetry component (the first and second principal components, respectively, of process and product IDE). We include the baseline controls: current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); log of current profit; current capital stock (Compustat ppeg, in logs); current employment (Compustat emp, in logs); product-weighted sales growth—the product-similarity-weighted average of contemporaneous sales growth across all other firms, where the weights are the pairwise product similarities from [Hoberg and Phillips \(2016\)](#); and aggregate firm-level innovation value, deflated by total assets. IDE measures, firm-level innovation value, and product-weighted sales growth are standardized across the entire sample. Fixed effects are sector-by-year and industry. Standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A. Revenue					
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}}$ (PC ₁)	-0.112*** (0.010)	-0.174*** (0.018)	-0.221*** (0.025)	-0.268*** (0.031)	-0.308*** (0.036)
IDE ^{tilt} (PC ₂)	-0.001 (0.002)	-0.001 (0.003)	-0.003 (0.004)	-0.006 (0.005)	-0.008 (0.007)
Observations	53,837	49,492	45,742	42,515	39,692
R ²	0.127	0.138	0.137	0.147	0.151
B. Market share					
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}}$ (PC ₁)	-0.108*** (0.011)	-0.165*** (0.019)	-0.211*** (0.026)	-0.264*** (0.032)	-0.318*** (0.038)
IDE ^{tilt} (PC ₂)	-0.000 (0.002)	-0.000 (0.003)	0.000 (0.005)	-0.002 (0.006)	-0.003 (0.007)
Observations	53,837	49,492	45,742	42,515	39,692
R ²	0.089	0.111	0.125	0.145	0.153
Fixed Effects:					
Year $\times$ Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓

TABLE 7: IDE and firm revenues and market share. This table shows the association between IDE measures and firm growth outcomes other than profit growth. Market share is the firm's revenue divided by the total revenue of all firms in the same 3-digit SIC industry and year. $\overline{\text{IDE}}$ (PC₁) is the composite displacement and IDE^{tilt} (PC₂) is the asymmetry component (the first and second principal components, respectively, of process and product IDE). We control for product-weighted sales growth (computed as the product-similarity-weighted average of growth in sales); aggregate innovation value of the focal firm in the current year; the focal firm's current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); current capital stock (Compustat ppeg, in logs); current employment (Compustat emp, in logs); and the current log level of the growth variable of interest if it is not already included. We standardize IDE, product-weighted sales growth, and firm-level innovation value across the entire sample. The fixed effects are sector-by-year plus industry. Standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A. Firm Employment					
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}} \text{ (PC}_1\text{)}$	-0.075*** (0.008)	-0.127*** (0.014)	-0.166*** (0.020)	-0.200*** (0.025)	-0.239*** (0.030)
$\text{IDE}^{\text{tilt}} \text{ (PC}_2\text{)}$	-0.001 (0.001)	-0.001 (0.003)	-0.002 (0.004)	-0.004 (0.005)	-0.005 (0.006)
Observations	53,822	49,506	45,749	42,561	39,761
R ²	0.070	0.080	0.083	0.087	0.094
B. Physical Capital					
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}} \text{ (PC}_1\text{)}$	-0.146*** (0.008)	-0.234*** (0.015)	-0.305*** (0.021)	-0.371*** (0.028)	-0.441*** (0.033)
$\text{IDE}^{\text{tilt}} \text{ (PC}_2\text{)}$	-0.003* (0.001)	-0.004 (0.003)	-0.006 (0.004)	-0.009* (0.005)	-0.010* (0.006)
Observations	53,930	49,602	45,847	42,625	38,173
R ²	0.148	0.153	0.162	0.165	0.170
C. Intangible Capital					
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}} \text{ (PC}_1\text{)}$	-0.110*** (0.005)	-0.192*** (0.010)	-0.264*** (0.016)	-0.323*** (0.021)	-0.369*** (0.026)
$\text{IDE}^{\text{tilt}} \text{ (PC}_2\text{)}$	-0.005*** (0.001)	-0.008*** (0.002)	-0.011*** (0.003)	-0.013*** (0.004)	-0.013** (0.005)
Observations	51,151	47,077	43,532	40,520	37,860
R ²	0.165	0.164	0.163	0.165	0.174
Fixed Effects:					
Year $\times$ Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓

TABLE 8: IDE and firm factor demand. This table shows the association between IDE measures and subsequent growth in the firm's factor demand: employment (Panel A), physical capital (Panel B), and intangible capital (Panel C). $\overline{\text{IDE}} \text{ (PC}_1\text{)}$ is the composite displacement and $\text{IDE}^{\text{tilt}} \text{ (PC}_2\text{)}$ is the asymmetry component (the first and second principal components, respectively, of process and product IDE). We control for product-weighted sales growth (computed as the product-similarity-weighted average of growth in sales); aggregate innovation value of the focal firm in the current year; the focal firm's current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); current capital stock (Compustat ppgt, in logs); current employment (Compustat emp, in logs); and the current log level of the growth variable of interest if it is not already included. We standardize IDE, product-weighted sales growth, and firm-level innovation value across the entire sample. The fixed effects are sector-by-year plus industry. Standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Dependent variable:	Standard controls			+Lagged (5yrs) Risk Score		
Specialist market tightening risk	splst risk $t + 1$ (1)	splst risk $t + 2$ (2)	splst risk $t + 3$ (3)	splst risk $t + 1$ (4)	splst risk $t + 2$ (5)	splst risk $t + 3$ (6)
$\overline{IDE}$ (PC ₁)	0.230*** (0.039)	0.235*** (0.042)	0.207*** (0.043)	0.275*** (0.071)	0.297*** (0.077)	0.275*** (0.082)
IDE^{tilt} (PC ₂)	-0.011 (0.008)	-0.012 (0.008)	-0.010 (0.009)	-0.001 (0.010)	-0.004 (0.011)	-0.007 (0.011)
Controls						
Year $\times$ Sector	✓	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓	✓
Lagged (5yrs) Risk Score				✓	✓	✓
Observations	46,013	39,834	35,058	26,373	23,114	20,381
R ²	0.202	0.205	0.211	0.298	0.288	0.280

TABLE 9: This table shows the association between IDE and future specialist labor market tightening risk over horizons of one to three years ahead. $\overline{IDE}$ (PC₁) is the composite displacement and IDE^{tilt} (PC₂) is the asymmetry component (the first and second principal components, respectively, of process and product IDE). Columns (1)–(3) include the standard controls listed below. Columns (4)–(6) additionally control for the specialist labor market tightening risk score of the focal firm five years ago, which requires a non-missing lagged score and reduces the number of observations from 46,013 to 26,373 at the one-year horizon. The standard controls are product-weighted sales growth (computed as the product-similarity-weighted average of growth in sales); aggregate innovation value of the focal firm in the current year; the focal firm’s current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); current capital stock (Compustat ppgt, in logs); current employment (Compustat emp, in logs); and current profit (Compustat sale minus cogs, in logs). We standardize the IDE measures, firm-level innovation value, product-weighted sales growth, and risk scores across the entire sample. The fixed effects are sector-by-year and industry. Standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Dependent variable:	Standard controls			+Lagged (5yrs) Risk Score		
Product obsolescence risk	ob risk $t + 1$ (1)	ob risk $t + 2$ (2)	ob risk $t + 3$ (3)	ob risk $t + 1$ (4)	ob risk $t + 2$ (5)	ob risk $t + 3$ (6)
$\overline{\text{IDE}}$ (PC ₁)	0.162*** (0.032)	0.122*** (0.035)	0.112*** (0.036)	0.260*** (0.065)	0.232*** (0.071)	0.255*** (0.073)
IDE ^{tilt} (PC ₂)	-0.001 (0.006)	-0.007 (0.007)	-0.002 (0.007)	0.001 (0.009)	-0.005 (0.010)	0.006 (0.010)
Controls						
Year $\times$ Sector	✓	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓	✓
Lagged (5yrs) Risk Score				✓	✓	✓
Observations	46,013	39,834	35,058	26,373	23,114	20,381
R ²	0.326	0.308	0.291	0.322	0.306	0.310

TABLE 10: This table shows the association between IDE and future product obsolescence risk over horizons of one to three years ahead. $\overline{\text{IDE}}$ (PC₁) is the composite displacement and IDE^{tilt} (PC₂) is the asymmetry component (the first and second principal components, respectively, of process and product IDE). Columns (1)–(3) include the standard controls listed below. Columns (4)–(6) additionally control for the obsolescence risk score of the focal firm five years ago, which requires a non-missing lagged score and reduces the number of observations from 46,013 to 26,373 at the one-year horizon. The standard controls are product-weighted sales growth (computed as the product-similarity-weighted average of growth in sales); aggregate innovation value of the focal firm in the current year; the focal firm's current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); log of current profit; current capital stock (Compustat ppeg, in logs); and current employment (Compustat emp, in logs). We standardize IDE measures, product-weighted sales growth, firm-level innovation, and obsolescence risk scores across the whole sample. The fixed effects are sector-by-year plus industry. Standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Horizon (years)				
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}} (\text{PC}_1)$	-0.122*** (0.011)	-0.190*** (0.019)	-0.232*** (0.025)	-0.283*** (0.031)	-0.332*** (0.037)
$\text{IDE}^{\text{tilt}} (\text{PC}_2)$	-0.002 (0.002)	-0.001 (0.003)	0.000 (0.005)	-0.005 (0.006)	-0.004 (0.007)
Product Market Similarity (log)	0.004 (0.010)	-0.009 (0.016)	-0.018 (0.021)	-0.013 (0.025)	0.008 (0.030)
$\overline{\text{IDE}} (\text{PC}_1) \times \text{Product Market Similarity (log)}$	-0.013*** (0.004)	-0.018*** (0.006)	-0.020** (0.009)	-0.030*** (0.009)	-0.039*** (0.011)
$\text{IDE}^{\text{tilt}} (\text{PC}_2) \times \text{Product Market Similarity (log)}$	0.000 (0.001)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	-0.001 (0.003)
Firm innovation value	0.028*** (0.002)	0.044*** (0.005)	0.057*** (0.006)	0.066*** (0.008)	0.075*** (0.009)
Year $\times$ Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓
Observations	49,519	45,496	42,079	39,154	36,575
R ²	0.134	0.152	0.148	0.147	0.153

TABLE 11: Profit growth, IDE, and Product Market Similarity. $\overline{\text{IDE}} (\text{PC}_1)$ is the composite displacement and $\text{IDE}^{\text{tilt}} (\text{PC}_2)$ is the asymmetry component (the first and second principal components, respectively, of process and product IDE). Product Market Similarity is the log of the sum of pairwise product-description cosine similarities between the focal firm and all other firms in the same year, following [Hoberg and Phillips \(2016\)](#); higher values indicate a more crowded product space. The interactions $\overline{\text{IDE}} \times \text{Product Market Similarity}$ and $\text{IDE}^{\text{tilt}} \times \text{Product Market Similarity}$ test whether the profit-growth associations of the composite displacement and the asymmetry component vary with how crowded the firm's product space is. Controls, fixed effects, and clustering are identical to the main analysis: firm-level innovation value; current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); log of current profit; current capital stock (Compustat ppeg, in logs); current employment (Compustat emp, in logs); and product-weighted sales growth. IDE components, Product Market Similarity, firm-level innovation value, and product-weighted sales growth are standardized across the whole sample. Fixed effects are sector-by-year and industry. Standard errors are clustered at the firm level. All firm-level variables are winsorized at the 1/99% level in each year. The sample period covers 1997 to 2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Forward Horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A					
$\overline{\text{IDE}} \text{ (PC}_1\text{)}$	-0.079*** (0.022)	-0.134*** (0.035)	-0.175*** (0.048)	-0.211*** (0.060)	-0.275*** (0.072)
$\text{IDE}^{\text{tilt}} \text{ (PC}_2\text{)}$	-0.005 (0.003)	-0.004 (0.005)	0.004 (0.008)	0.002 (0.009)	0.007 (0.011)
Firm innovation value	0.028*** (0.003)	0.045*** (0.005)	0.059*** (0.008)	0.069*** (0.009)	0.078*** (0.011)
Observations	14,841	13,899	13,013	12,251	11,610
R ²	0.161	0.187	0.196	0.198	0.204
Panel B:					
$\overline{\text{IDE}}$	-0.078*** (0.022)	-0.133*** (0.035)	-0.174*** (0.047)	-0.209*** (0.059)	-0.273*** (0.072)
IDE^{tilt}	-0.005 (0.003)	-0.004 (0.005)	0.004 (0.008)	0.002 (0.009)	0.007 (0.011)
Technological obsolescence	-0.008** (0.004)	-0.008 (0.006)	-0.007 (0.008)	-0.008 (0.010)	-0.009 (0.012)
Firm innovation value	0.028*** (0.003)	0.045*** (0.005)	0.059*** (0.008)	0.069*** (0.009)	0.078*** (0.011)
Observations	14,841	13,899	13,013	12,251	11,610
R ²	0.161	0.187	0.196	0.199	0.204
Fixed Effects:					
Industry	✓	✓	✓	✓	✓
Year $\times$ Sector	✓	✓	✓	✓	✓

TABLE 12: This table shows the association between IDE and profit growth over the next five years (one to five years ahead). Both Panel A and Panel B are estimated on the subsample of patenting firms over the sample period 1997–2016, the overlap window for which the [Ma \(2025\)](#) citation-based technological obsolescence measure is available. Panel A reports the baseline IDE specification on this subsample; Panel B additionally controls for the technological obsolescence measure. $\overline{\text{IDE}}$ (PC₁) is the composite displacement and IDE^{tilt} (PC₂) is the asymmetry component (the first and second principal components, respectively, of process and product IDE). We include full controls: product-weighted sales growth (sales growth weighted by product similarity); aggregate firm-level innovation value; current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); log of current profit; current capital stock (Compustat ppgt, in logs); current employment (Compustat emp, in logs). We standardize IDE measures, technological obsolescence, product-weighted sales growth, and firm-level innovation value across the whole sample. Fixed effects are sector-by-year plus industry. Standard errors are clustered at the firm level. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Industry Growth (Compustat)	Forward horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A: Revenue per employee, $\Delta_k \log(\sum \text{sale} / \sum \text{emp})$					
$\overline{\text{IDE}}$ (PC ₁)	0.038 (0.045)	0.031 (0.058)	0.009 (0.066)	−0.022 (0.078)	−0.021 (0.093)
IDE^{tilt} (PC ₂)	0.007 (0.011)	−0.007 (0.018)	−0.012 (0.020)	−0.006 (0.025)	−0.017 (0.028)
Observations	3,082	3,082	3,082	3,082	3,082
R ²	0.074	0.091	0.104	0.114	0.135
Panel B: Gross profit per employee, $\Delta_k \log((\sum \text{sale} - \sum \text{cogs}) / \sum \text{emp})$					
$\overline{\text{IDE}}$ (PC ₁)	0.067 (0.060)	0.036 (0.074)	−0.016 (0.100)	−0.048 (0.109)	−0.002 (0.127)
IDE^{tilt} (PC ₂)	0.024 (0.017)	0.015 (0.026)	0.007 (0.029)	0.017 (0.035)	0.008 (0.038)
Observations	3,075	3,072	3,071	3,071	3,071
R ²	0.046	0.067	0.079	0.082	0.091

TABLE 13: The table reports two definitions of industry-level labor productivity, each formed as a sum-of-sums ratio across all Compustat firms in the industry-year cell. Panel A reports revenue per employee, $\log(\sum_f \text{sale}_f / \sum_f \text{emp}_f)$. Panel B reports gross profit per employee, $\log((\sum_f \text{sale}_f - \sum_f \text{cogs}_f) / \sum_f \text{emp}_f)$. The dependent variable in each panel is the k -year forward log change of the corresponding industry-level ratio. The unit of observation is an industry-year defined by the 3-digit SIC code, with a minimum of five IDE-valid firms per cell. The common component, $\overline{\text{IDE}}$, and the tilt component, IDE^{tilt} , are the first and second principal components of the asset-weighted industry means of $\text{IDE}^{\text{process}}$ and $\text{IDE}^{\text{product}}$. Each is converted to its empirical percentile rank in $[0, 1]$ across the full industry-year panel before entering the regression, with ties averaged; a unit change in the regressor corresponds to moving from the 0th to the 100th percentile. The current-period industry log level of the productivity ratio enters as a control. We include year fixed effects in the specification. Standard errors are clustered at the industry level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Industry Growth (BEA)	Forward horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$\overline{\text{IDE}}$ (PC ₁)	+0.018 (0.020)	+0.048 (0.035)	+0.072 (0.049)	+0.061 (0.064)	+0.088 (0.076)
IDE^{tilt} (PC ₂)	−0.007 (0.006)	−0.007 (0.010)	−0.015 (0.013)	−0.021 (0.016)	−0.028 (0.019)
Observations	1,934	1,809	1,693	1,592	1,496
R^2	0.282	0.403	0.500	0.579	0.647

TABLE 14: IDE and forward industry-level BEA labor productivity growth. The unit of observation is an industry-year, where the industry is the 4-digit NAICS code obtained by truncating each Compustat firm's NAICS string to the first four characters; codes that do not appear in the BEA file at this depth are dropped. The dependent variable is forward log-growth in the BEA labor productivity index. The common component, $\overline{\text{IDE}}$, and the tilt component, IDE^{tilt} , are the first and second principal components of the asset-weighted industry means of $\text{IDE}^{\text{process}}$ and $\text{IDE}^{\text{product}}$. Both regressors enter as empirical percentile ranks in $[0, 1]$ across the industry-year panel, with ties averaged; a unit change corresponds to moving from the 0th to the 100th percentile. The current-period log of the dependent-variable index enters as a level control. Year fixed effects are absorbed. Standard errors are clustered at the 4-digit NAICS industry level. The sample requires at least five IDE-valid Compustat firms per industry-year cell. Firm-level inputs, $\text{IDE}^{\text{process}}$, $\text{IDE}^{\text{product}}$, A^f , and at , are winsorized at 1% and 99% within each year before aggregation. The sample period covers 1997 to 2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Industry Reallocation: Dispersion in Firm Growth Rates	Forward horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A: Growth in Revenue per employee, $\Delta_k \log(\text{sale}/\text{emp})_f$					
$\overline{\text{IDE}}$ (PC ₁)	0.571*** (0.172)	0.649*** (0.232)	0.659** (0.277)	0.787** (0.324)	0.671* (0.354)
IDE^{tilt} (PC ₂)	0.030 (0.029)	0.008 (0.036)	−0.001 (0.041)	−0.004 (0.046)	−0.005 (0.051)
Observations	2,871	2,730	2,589	2,457	2,334
R^2	0.285	0.308	0.322	0.339	0.349
Panel B: Growth in Gross profit per employee, $\Delta_k \log((\text{sale} - \text{cogs})/\text{emp})_f$					
$\overline{\text{IDE}}$ (PC ₁)	0.399*** (0.128)	0.355** (0.164)	0.331* (0.193)	0.377* (0.219)	0.403* (0.244)
IDE^{tilt} (PC ₂)	0.035 (0.033)	0.013 (0.045)	−0.011 (0.052)	−0.001 (0.057)	0.032 (0.062)
Observations	2,820	2,676	2,535	2,409	2,287
R^2	0.198	0.193	0.196	0.201	0.209

TABLE 15: This table reports the association between industry-level IDE and industry-level dispersion in firm growth rates over forward horizons of one to five years. Panels A and B report the within-industry P90–P10 spread of firm-level k -year log growth in revenue per employee, $\Delta_k \log(\text{sale}/\text{emp})_f$, and gross profit per employee, $\Delta_k \log((\text{sale} - \text{cogs})/\text{emp})_f$, respectively. The common component, $\overline{\text{IDE}}$ (PC₁), and the tilt component, IDE^{tilt} (PC₂), are the first and second principal components of the asset-weighted industry means of process and product IDE. Both regressors enter as empirical percentile ranks in $[0, 1]$ across the industry-year panel, with ties averaged; a coefficient is associated with moving from the 0th to the 100th percentile of the regressor. Panels A and B control for the contemporaneous within-industry spread of the corresponding level variable. Year fixed effects are absorbed. Standard errors are clustered at the industry level. All firm-level variables are winsorized at the 1/99% level in each year. The sample period covers 1997 to 2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Forward Horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
<i>Panel A: baseline specification</i>					
$\overline{\text{IDE}}$ (PC ₁)	−0.106*** (0.011)	−0.173*** (0.018)	−0.211*** (0.024)	−0.250*** (0.029)	−0.290*** (0.035)
IDE^{tilt} (PC ₂)	−0.003 (0.002)	−0.002 (0.003)	−0.000 (0.004)	−0.006 (0.005)	−0.005 (0.006)
Firm innovation value	0.027*** (0.002)	0.043*** (0.004)	0.054*** (0.006)	0.063*** (0.008)	0.073*** (0.009)
Randomized-timing placebos	—	—	—	—	—
Observations	50,028	45,970	42,522	39,567	36,963
R ²	0.134	0.151	0.146	0.146	0.152
<i>Panel B: adding the randomized-timing placebos as controls</i>					
$\overline{\text{IDE}}$ (PC ₁)	−0.118*** (0.011)	−0.177*** (0.018)	−0.209*** (0.024)	−0.240*** (0.029)	−0.276*** (0.034)
IDE^{tilt} (PC ₂)	−0.004* (0.002)	−0.003 (0.003)	−0.001 (0.004)	−0.006 (0.005)	−0.004 (0.006)
Firm innovation value	0.027*** (0.002)	0.043*** (0.004)	0.054*** (0.006)	0.063*** (0.008)	0.073*** (0.009)
Randomized-timing placebos	✓	✓	✓	✓	✓
Observations	50,028	45,970	42,522	39,567	36,963
R ²	0.134	0.151	0.146	0.146	0.152
Year × Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓

TABLE 16: Placebo 1: randomized technology-stack timing. Panel A reports the baseline profit-growth specification. Panel B adds the randomized-timing placebo measures as controls; because the placebo terms are a nuisance control rather than an object of interest, their presence is marked with a checkmark and their coefficients are not shown. The placebo measures are the composite and asymmetry components formed by applying the paper’s principal-component convention to process and product IDE computed after randomly permuting, within each firm, the year ordering of the five-year technology-stack inputs, holding the innovation side fixed; process and product channels are shuffled separately before the components are formed. The comparison of interest is between the correctly timed $\overline{\text{IDE}}$ and IDE^{tilt} rows across the two panels: under the null that within-firm timing is irrelevant, the placebo analogs would absorb the predictive content and the Panel B coefficients would collapse. Both panels include the full set of baseline controls: product-weighted sales growth (sales growth weighted by product similarity); aggregate firm-level innovation value (deflated by total assets); current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); log of current profit; current capital stock (Compustat ppeg, in logs); current employment (Compustat emp, in logs). We standardize IDE measures, product-weighted sales growth, and firm-level innovation value across the entire sample. Fixed effects are sector-by-year and industry. Standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Forward Horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A: contemporaneous IDE (baseline)					
$\overline{\text{IDE}} (\text{PC}_1)$	-0.106^{***} (0.011)	-0.173^{***} (0.018)	-0.211^{***} (0.024)	-0.250^{***} (0.029)	-0.290^{***} (0.035)
$\text{IDE}^{\text{tilt}} (\text{PC}_2)$	-0.003 (0.002)	-0.002 (0.003)	-0.000 (0.004)	-0.006 (0.005)	-0.005 (0.006)
Observations	50,028	45,970	42,522	39,567	36,963
R ²	0.134	0.151	0.146	0.146	0.152
Panel B: IDE dated $t + 5$					
$\overline{\text{IDE}} (\text{PC}_1), t + 5$	-0.037^* (0.022)	-0.059 (0.037)	-0.062 (0.053)	-0.085 (0.064)	-0.116 (0.073)
$\text{IDE}^{\text{tilt}} (\text{PC}_2), t + 5$	0.000 (0.003)	-0.002 (0.004)	0.001 (0.006)	-0.001 (0.007)	-0.003 (0.008)
Observations	28,832	28,730	28,692	28,638	28,664
R ²	0.139	0.165	0.162	0.160	0.155
Panel C: IDE dated $t + 7$					
$\overline{\text{IDE}} (\text{PC}_1), t + 7$	-0.038 (0.028)	-0.072 (0.047)	-0.080 (0.064)	-0.109 (0.078)	-0.138 (0.092)
$\text{IDE}^{\text{tilt}} (\text{PC}_2), t + 7$	-0.001 (0.003)	-0.002 (0.005)	-0.001 (0.007)	-0.005 (0.008)	-0.003 (0.009)
Observations	22,250	22,181	22,125	22,076	22,066
R ²	0.151	0.171	0.173	0.179	0.178

TABLE 17: Look-ahead placebo. Panel A reproduces the baseline profit-growth regression with contemporaneous IDE. Panels B and C replace it with the IDE measured five and seven years later, holding the controls, the outcome window and the fixed effects at t . Because IDE at date s is constructed from a five-year look-back over filings $[s - 4, s]$, the measure at $t + 5$ is the first vintage built entirely from filings dated after t . The sample is the baseline estimation sample throughout, so the panels differ only in the vintage of IDE. Controls, fixed effects and clustering are as in Table 6. The data are from 1997 to 2019. $***p < 0.01$, $**p < 0.05$, $*p < 0.10$.

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Forward Horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$\overline{\text{IDE}}$ (PC ₁), unweighted	−0.043*** (0.004)	−0.068*** (0.007)	−0.084*** (0.009)	−0.101*** (0.011)	−0.119*** (0.014)
IDE^{tilt} (PC ₂), unweighted	−0.002 (0.002)	−0.001 (0.003)	0.002 (0.004)	−0.002 (0.005)	−0.001 (0.006)
Firm innovation value	0.026*** (0.002)	0.042*** (0.004)	0.053*** (0.006)	0.062*** (0.008)	0.072*** (0.009)
Year $\times$ Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓
Observations	50,028	45,970	42,522	39,567	36,963
R ²	0.134	0.152	0.147	0.147	0.153

TABLE 18: This table shows the association between the unweighted IDE measures and profit growth over the next five years (one to five years ahead). $\overline{\text{IDE}}$ (PC₁) is the composite displacement and IDE^{tilt} (PC₂) is the asymmetry component (the first and second principal components, respectively, of process and product IDE); both are computed here without weighting other firms' innovations by firm-level innovation value. We include the full set of controls: product-weighted sales growth (sales growth weighted by product similarity); aggregate firm-level innovation value (deflated by total assets); current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); log of current profit; current capital stock (Compustat ppegst, in logs); current employment (Compustat emp, in logs). We standardize IDE measures, firm-level innovation value, and product-weighted sales growth across the entire sample. Fixed effects are sector-by-year and industry. Standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Forward Horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$\overline{\text{IDE}}$ (PC ₁)	-0.102*** (0.015)	-0.148*** (0.025)	-0.180*** (0.034)	-0.219*** (0.041)	-0.240*** (0.050)
IDE ^{tilt} (PC ₂)	-0.000 (0.003)	0.003 (0.005)	0.008 (0.006)	0.006 (0.008)	0.007 (0.009)
Firm innovation value	0.027*** (0.003)	0.043*** (0.005)	0.058*** (0.007)	0.067*** (0.009)	0.078*** (0.011)
Year $\times$ Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓
Observations	29,418	26,770	24,559	22,610	20,947
R ²	0.156	0.177	0.173	0.174	0.182

TABLE 19: This table shows the association between residual IDE measures and profit growth over the next five years (one to five years ahead). $\overline{\text{IDE}}$ (PC₁) is the composite displacement and IDE^{tilt} (PC₂) is the asymmetry component (the first and second principal components, respectively, of process and product IDE). We include the full set of standardized controls: firm age; Product Market Similarity (the total product-similarity sum); product-weighted sales growth (sales growth weighted by product similarity); firm-level innovation value, deflated by total assets; current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); log of current profit; current capital stock (Compustat ppeg, in logs); current employment (Compustat emp, in logs); and variable cost, defined as the ratio of Compustat cogs to sale. IDE measures, Product Market Similarity, product-weighted sales growth, and firm-level innovation values are standardized across the whole sample. Fixed effects are sector-by-year and industry, and standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Forward Horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A: $\overline{\text{IDE}}$ and IDE^{tilt} , high (2SD above mean)					
$\overline{\text{IDE}}$, high (PC ₁)	−0.008** (0.003)	−0.022*** (0.006)	−0.029*** (0.008)	−0.036*** (0.009)	−0.047*** (0.011)
IDE^{tilt} , high (PC ₂)	−0.004** (0.002)	−0.006 (0.004)	−0.004 (0.005)	−0.005 (0.006)	0.001 (0.007)
Observations	50,028	45,970	42,522	39,567	36,963
R ²	0.131	0.149	0.143	0.143	0.149
Panel B: $\overline{\text{IDE}}$ and IDE^{tilt} , moderate (1SD; controlling for high)					
$\overline{\text{IDE}}$, moderate (PC ₁)	−0.031*** (0.005)	−0.042*** (0.009)	−0.061*** (0.012)	−0.068*** (0.014)	−0.065*** (0.017)
IDE^{tilt} , moderate (PC ₂)	−0.000 (0.003)	0.004 (0.004)	0.007 (0.006)	0.003 (0.007)	−0.002 (0.008)
$\overline{\text{IDE}}$, high (PC ₁)	0.017*** (0.006)	0.012 (0.010)	0.021* (0.013)	0.020 (0.015)	0.005 (0.018)
IDE^{tilt} , high (PC ₂)	−0.004 (0.003)	−0.010* (0.005)	−0.010 (0.007)	−0.008 (0.008)	0.002 (0.009)
Observations	50,028	45,970	42,522	39,567	36,963
R ²	0.132	0.149	0.145	0.144	0.150
Panel C: Baseline $\overline{\text{IDE}}$ and IDE^{tilt} (controlling for moderate)					
$\overline{\text{IDE}}$ (PC ₁)	−0.155*** (0.016)	−0.228*** (0.027)	−0.247*** (0.035)	−0.291*** (0.043)	−0.335*** (0.051)
IDE^{tilt} (PC ₂)	−0.002 (0.003)	−0.004 (0.005)	−0.003 (0.006)	−0.012 (0.008)	−0.013 (0.009)
$\overline{\text{IDE}}$, moderate (PC ₁)	0.019*** (0.005)	0.020*** (0.008)	0.012 (0.011)	0.013 (0.013)	0.013 (0.015)
IDE^{tilt} , moderate (PC ₂)	−0.001 (0.003)	0.002 (0.005)	0.004 (0.006)	0.008 (0.007)	0.011 (0.009)
Observations	50,028	45,970	42,522	39,567	36,963
R ²	0.134	0.152	0.147	0.146	0.152

TABLE 20: This table reports the association between future profit growth and $\overline{\text{IDE}}$ (PC₁, the composite displacement) and IDE^{tilt} (PC₂, the asymmetry component)—the first and second principal components of process and product IDE—measured at different similarity thresholds. The *high* versions of these factors are the first two principal components of the process and product IDE constructed by aggregating only over firms whose innovation similarity to the focal firm is at least two standard deviations above the mean similarity across all firm pairs and years. The *moderate* versions are defined analogously using a one-standard-deviation threshold. Panel A includes the high $\overline{\text{IDE}}$ and high IDE^{tilt} . Panel B adds the moderate $\overline{\text{IDE}}$ and moderate IDE^{tilt} while controlling for the high (two-standard-deviation) versions. Panel C includes the baseline $\overline{\text{IDE}}$ and IDE^{tilt} while controlling for the moderate (one-standard-deviation) versions. In all specifications we control for product-weighted sales growth; aggregate firm-level innovation value; current profitability; log current profit; current capital stock, current employment. We standardize all IDE measures, product-weighted sales growth, and firm-level innovation value across the whole sample. Fixed effects are sector-by-year and industry. Standard errors are clustered at the firm level. The sample period is 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix

- The first two subsections of Section **A** describe the dataset and variable definitions, and the third describes industry-level aggregation and percentile ranks. Subsection **A.4** reviews the literature on processing 10-K documents. Subsection **A.5** gives the prompt designs, with the SemanticAxis sentence pairs in its two subsubsections; Subsections **A.6** and **A.7** give example summaries. Subsection **A.8** shows alternative prompts that produce nearly identical outputs.
- Section **B** uses Microsoft as an example to showcase the evolution of technologies captured by our summaries.
- Section **C** examines IDE interacted with firm characteristics.
- Section **D** contains the full theoretical model motivating the process and product channels of displacement.
- Section **E** includes additional analyses where we analyze the association between IDE and the focal firm's profitability growth and benchmark the magnitude of our main results against R&D and capital expenditures.

A Data Construction Appendix

This appendix collects the data sources, the construction details for variables that are not readily available from Compustat, and a small number of supporting exhibits.

A.1 Dataset construction

We pull firm characteristics from the CRSP/Compustat Merged database (Compustat North America, Fundamentals Annual, linked to CRSP by PERMNO) through fiscal year 2024, the Consumer Price Index and the equipment deflator from the Federal Reserve Bank of St. Louis, and firm- and industry-level innovation values from [the KPSS data repository](#). The datasets are merged on year and the unique firm identifier PERMNO. Throughout, *year* denotes the Compustat fiscal year, and all by-year operations — winsorization, standardization and the year fixed effects — use that definition.

Six firm-level outcomes enter the analysis—profit growth, physical capital stock growth, intangible capital stock growth, revenue growth, employment growth, and market share growth—each in real terms. We deflate physical capital stock by the equipment deflator and all other outcomes by the Consumer Price Index.

Industry classifications follow the [Fama-French](#) 30 industry definitions. Financials (SIC 6000–6799) and utilities (SIC 4900–4949) are excluded by an explicit screen at the 10-K collection stage, so those two categories are absent from the firm panel entirely rather than present but unscored; 28 sectors appear in the IDE-valid panel: Personal and Business Services; Business Equipment; Healthcare, Medical Equipment, and Pharmaceutical Products; Retail; Petroleum and Natural Gas; Wholesale; Communication; Everything Else; Construction and Construction Materials; Fabricated Products and Machinery; Transportation; Restaurants, Hotels, and Motels; Recreation; Consumer Goods; Food Products; Chemicals; Automobiles and Trucks; Business Supplies and Shipping Containers; Steel Works; Precious Metals, Non-Metallic, and Industrial Metal Mining; Apparel; Electrical Equipment; Printing and Publishing; Aircraft, Ships, and Railroad Equipment; Textiles; Beer and Liquor; Tobacco Products; and Coal. Figure 3 reports the count of unique firms in each sector over the IDE-valid panel.

The merged firm-level panel covers 1997 to 2019 and contains 70,236 firm-year observations on 8,753 unique firms with valid IDE scores.

A.2 Variable dictionary

- Profit: total sales minus cost of goods sold. Deflated by the consumer price index.
- Capital stock (physical) is defined as Property, Plant, and Equipment (Gross) divided by the equipment deflator.
- Intangible capital is computed using the perpetual-inventory method that capitalizes past R&D and a fraction of past SG&A with sector-specific depreciation rates, as in [L. Eisfeldt, T. Kim and Papanikolaou \(2022\)](#).
- Profitability is defined as profit divided by total assets.
- Asset growth is defined as the log of total assets in the current year divided by the total assets in the previous year.
- PERMNO: unique firm identifier used to merge datasets.

- Industry (code): The 3-digit SIC industry code.
- Sectors are defined by the [Fama-French](#) 30 industry categories, of which 28 appear in our panel (Finance and Utilities are absent).
- Firm-level innovation values are based on [the KPSS data repository](#) where innovations are aggregated at the firm-year level and scaled by the firm's total assets in that year.
- Product Market Similarity (PMS) is defined as the sum of product embedding cosine similarity with respect to each firm in each year.
- Product-weighted sales growth is defined as the sum of growth in sales from $t - 1$ to t weighted by the product embedding cosine similarity with respect to each firm in each year.
- Variable cost is defined as the cost of goods sold divided by revenue.
- Market share is defined by the revenue of the firm divided by the total revenue of all firms in the same 3-digit SIC industry.
- Age is computed as the number of years since the IPO of a firm based on [the Jay Ritter IPO data repository](#).

A.3 Industry-level aggregation and percentile ranks

This subsection collects the construction details that support the industry-level evidence in Section 6. The body of that section keeps a lean summary; the rules below are the ones we actually impose.

Sample and aggregation. Industries are 3-digit SIC codes from Compustat. For each industry-year (i, t) we form asset-weighted means of $\text{IDE}_{f,t}^{\text{process}}$ and $\text{IDE}_{f,t}^{\text{product}}$ across firms in industry i at date t for which the channel-specific IDE measure and Compustat total assets at are both non-missing and strictly positive; weights are $\text{at}_{f,t}$. We retain industry-year cells with at least five IDE-valid firms and use the resulting panel from 1997 to 2019. The dispersion regressions in Section 6.2 additionally require five firms in the cell with non-missing realizations of the underlying spread variable.

Productivity ratios. For each firm-year, gross profit per worker and revenue per worker are gross profit (Compustat `sale` minus `cogs`) and sales divided by employment, both in natural logs. For each horizon $k \in \{1, \dots, 5\}$ we take k -year forward changes in these log ratios and retain the contemporaneous log levels as controls for the spread regressions. We winsorize the resulting firm-level series at the first and ninety-ninth percentiles within each year, separately for each variable.

Winsorization before aggregation. Before forming asset-weighted industry means we winsorize firm-level $\text{IDE}_{f,t}^{\text{process}}$, $\text{IDE}_{f,t}^{\text{product}}$, the firm-level innovation value $A_{f,t}^f$, and Compustat `at` at the first and ninety-ninth percentiles on each tail within each year. The same rule applies to the firm-level inputs of every regression table; the one exception is the summary-statistics table, whose inputs are winsorized on the pooled panel.

Industry-level principal components. We re-extract the first two principal components on the industry-year panel rather than reusing the firm-year loadings from Section 2.5. The PCA is taken on the correlation matrix of the two asset-weighted industry means—each series is centered and scaled before the eigen-decomposition. Two sign conventions match the firm-level construction: if both channel-specific means do not load positively on the first component, we flip its sign so both loadings are non-negative; if the process-IDE mean does not load positively on the second component, we flip its sign so the process loading is positive. Both industry-level aggregates are right-skewed, so we map each to empirical percentile ranks in $[0, 1]$ across the full surviving industry-year panel, with ties averaged to the mid-rank. We denote the rank-transformed series by $\overline{\text{IDE}}_{i,t}$ and $\text{IDE}_{i,t}^{\text{tilt}}$; a one-unit change in either is the move from the 0th to the 100th percentile.

Asset-weighted aggregate of own-firm innovation value. We construct $\overline{A}_{i,t}^f$ as the asset-weighted industry mean of $A_{f,t}^f$ on the same sample, then percentile-rank it in $[0, 1]$ with ties averaged the same way. We include $\overline{A}_{i,t}^f$ only as a sanity check in the average-productivity regressions in Section 6.1 and do not enter it in the dispersion or exit specifications.

BEA exception. Table 14 matches Compustat industry-year aggregates of $\overline{\text{IDE}}$, IDE^{tilt} , and $\overline{A}^f$ to the BEA published 4-digit NAICS labor-productivity index. For that exercise industries are 4-digit NAICS codes obtained by truncating each Compustat firm’s NAICS string to the first four characters; codes that do not appear in the BEA file at this depth are dropped, and standard errors are clustered at the 4-digit NAICS industry rather than the 3-digit SIC industry used elsewhere.

A.4 Papers using 10-Ks

A large body of literature shows that 10-K disclosures contain economically meaningful information about firms’ fundamentals and expectations. For example, annual-report readability is related to earnings persistence Li (2008); investors underreact more to 10-K information when filings are more complex You and Zhang (2009); forward-looking statements in the MD&A predict future earnings Li (2010); and textual changes in periodic reports predict future returns and fundamentals Cohen, Malloy and Nguyen (2020). At the same time, this literature also highlights several well-known measurement challenges. Generic sentiment dictionaries can misclassify financial language Loughran and McDonald (2011); conventional readability measures such as the Fog index are imperfect in business disclosures Loughran and McDonald (2014); and 10-Ks have become longer, more boilerplate, and more persistent over time Dyer, Lang and Stice-Lawrence (2017).⁴ More broadly, Loughran and McDonald (2016) emphasize that textual analysis in accounting and finance is inherently noisy and sensitive to implementation choices.

To overcome these challenges in information extraction from 10-Ks we use LLMs (GPT4o-mini and GPT5.1) to extract information strictly related to our application — the firm’s process and product technologies. The exact prompts are shown below.

A.5 10-K prompts

A.5.1 Prompt to extract technology descriptions

Stage 1:

⁴We directly use the full dataset by Loughran and McDonald (2014) as the source of 10-K from 1993 to 2000 and do not apply truncation when processing.

Prompt (10-K Technology Extraction). You are an economist analyzing a firm's 10-K filing to extract and summarize explicit mentions of technologies (hardware, software, platforms, systems, processes, and R&D projects).

Goals.

1. Identify each technology mentioned; be specific.
2. Classify each technology into exactly one of:
 - "input_only": used only to support internal operations / production / R&D (not sold or licensed), or
 - "product": offered to customers as a product or service (even if also used internally), or
 - "unclear": mentioned, but internal-only vs. customer-facing is not explicitly stated.

Classification rule (non-negotiable). If a technology is used both internally and offered externally, you must classify it as "product" (there is no "both" category). You may note internal use only if the filing explicitly states it.

Hard constraints (must follow).

- Use **only** information explicitly stated in the 10-K; do **not** infer or speculate.
- Attach a **section + page reference to every factual point** (e.g., (Business, p. 12)).
- If a page number cannot be determined, use (section_name, page_number_unknown) and label it as unknown.
- Organize the answer into **exactly two sections**: (I) inputs; (II) products/services.

Scope (search only these 10-K sections).

1. Business
2. Risk Factors
3. Properties
4. MD&A
5. Any R&D-specific sections

Ignore boilerplate cover pages, table of contents, and signatures unless they contain substantive technology descriptions.

Extraction Guidelines

1. Identify all explicit technology mentions.

- Capture every explicit mention of technologies, tools, systems, platforms, hardware/software, proprietary processes, and R&D projects/initiatives.
- Include internal operations technologies (e.g., manufacturing, logistics, internal IT/data systems), R&D support technologies, and technologies embedded in or constituting customer offerings.
- If the filing clearly refers to a family of related technologies as a single concept (e.g., "our cloud-based analytics platform"), treat it as one item.
- For each item, record at least one source location (Section, p. X-Y).

2. For each technology, extract the following (each bullet must end with a reference).

- **Name/description:** the exact term(s) used in the filing, or a concise description using only filing language. (Section, p. X)
- **Classification (required):** "input_only" / "product" / "unclear" with a one-sentence justification quoting or closely paraphrasing the filing. (Section, p. X)
- **Function / use case:** how it is used; cite each distinct function statement separately. (Section, p. X)
- **Key features / capabilities:** only those explicitly stated; cite each feature (or tightly related group) separately. (Section, p. X)
- **Partnerships / suppliers:** any named third parties and their explicitly stated role; cite each partner/role.

(Section, p. X)

3. Relationships between technologies.

- Note only explicitly stated integrations or interactions; cite each relationship description. (Section, p. X)

Required Output Structure (Content)

I. Technologies Used as Inputs (internal only).

- Include all items classified as "input_only".
- For each item: name/description; classification + justification; functions; features/capabilities (if any); partnerships/suppliers (if any); and a list of all section/page locations.
- End with a brief synthesis paragraph; every sentence must be supported by prior citations.

II. Technologies Offered as Products or Services (customer-facing).

- Include all items classified as "product" (including those also used internally).
- For each item: name/description; classification + justification; whether also used internally (only if explicitly stated); customer-facing functions; any explicitly stated business rationale, risks, and partnerships; and a list of all section/page locations.
- End with a brief synthesis paragraph; every sentence must be supported by prior citations.

Strict output format. Return only a valid JSON array of exactly two strings and nothing else:

```
[  
  "<FULL Section I text incl. heading + synthesis>",  
  "<FULL Section II text incl. heading + synthesis>"  
]
```

Stage 2:

Input-only prompt. *Role:* You are an economist.

Objective: Read the provided **summary** carefully and rewrite it in simplified form to extract only technologies that are **strictly inputs**—used internally only for operations, production support, internal IT/information systems, or R&D support tools.

Rules (strict):

- Use only information explicitly contained in the summary; do not use outside knowledge and do not speculate.
- Keep only technologies that are unambiguously internal-only. If a technology is ambiguous, dual-use, or could be customer-facing, exclude it.
- Exclude customer-facing products/services, platforms sold/licensed to customers, and market-facing offerings.
- If the summary contains citations (e.g., section/page references), retain them on the extracted statements.
- If no unambiguous input technologies are mentioned, output exactly:

No unambiguous input-only technologies found in the provided summary.

Output format: Plain text only; prefer bullet points; keep wording close to the summary; no extra commentary.

Output-only prompt. *Role:* You are an economist.

Objective: Read the provided **summary** carefully and rewrite it in simplified form to extract only technologies that are **strictly outputs**—customer-facing products or services the firm sells, licenses, or delivers (including licensed technology, platforms, SaaS, devices, and other offerings).

Rules (strict):

- Use only information explicitly contained in the summary; do not use outside knowledge and do not speculate.
- Keep only technologies that are unambiguously customer-facing. If a technology is ambiguous, dual-use, or could be internal-only, exclude it.
- Exclude internal-only operational tools, internal IT/data systems used only for operations, and R&D support tools not offered externally.
- If the summary contains citations (e.g., section/page references), retain them on the extracted statements.
- If no unambiguous output technologies are mentioned, output exactly:

No unambiguous output technologies found in the provided summary.

Output format: Plain text only; prefer bullet points; keep wording close to the summary; no extra commentary.

A.5.2 Prompt to extract discussions of obsolescence risk

Stage 1:

Obsolescence-risk severity prompt (10-K). **Role:** You are a financial economist.

Objective: Analyze the provided 10-K text to extract explicit evidence about the **severity of obsolescence risk** (the risk that technologies become outdated). Use only what is explicitly stated in the 10-K; do not use outside knowledge and do not speculate.

Required exposure types (cover both if present).

- **(A) Internal operating capabilities:** internal technologies/capabilities used for operations, production, or research (equipment, facilities, tooling, systems/IT, processes/methods/know-how, automation, service-delivery methods, supply chain/logistics systems, controls/procedures, workforce training/retention, and capex/modernization needed to keep pace).
- **(B) Customer offerings:** customer-facing products/services becoming outdated or losing demand due to technological change (substitution, uncompetitiveness, shortening life cycles, and customer switching driven by new technology).

Internal-Capability Triggers (Low Threshold in Step 1)

Treat a statement as internal-capability-related (eligible for Section I) if it contains:

- any internal-object cue (e.g., equipment, systems/IT, processes, training, capex/modernization), and
- any tech-change / keep-pace cue (e.g., rapid change, innovation, modernize/upgrade/replace/invest/implement new systems), and/or
- any consequence / severity cue (e.g., materially adversely affect, significant investment, disruption, increased costs, competitive disadvantage).

Do not require the word “obsolete”: many filings express input obsolescence as “need to invest/upgrade/modernize/keep pace.”

Mandatory Workflow (Do in Order)

1. **Step 1 (Internal-capability pass):** Extract all explicit severity statements tied to internal operating capabilities. Do not proceed until you have extracted all internal-capability instances you can find.
2. **Step 2 (Customer-offering pass):** Extract all explicit severity statements tied to customer-facing offerings.
3. **Step 3 (Unclear pass):** Collect remaining tech-change/obsolescence severity statements that do not clearly map to (A) or (B).
4. **Step 4 (Input recall audit):** Re-check every quote considered in Steps 2–3. If it mentions internal triggers and the mechanism is internal upgrading/capability maintenance, move it into Step 1 evidence (Section I) while also keeping the original entry in Step 2 or Step 3 evidence.

Tie-Breaker (Who Is Harmed?)

- If the harm is the firm needing to upgrade/modernize internal capabilities to remain competitive → **Internal (Step 1)**.
- If the harm is customers switching away / offerings losing demand → **Offering (Step 2)**.
- Otherwise → **Unclear (Step 3)**.

Extraction Rules (Strict)

1. **Exact quotes only:** quote the smallest complete sentence(s) containing severity language; no paraphrase.
2. **Section + page for every quote:** end each quote with (Section, p. X) or (Section, p. X-Y); if unknown use (Section, page_number_unknown).
3. **Split mixed paragraphs:** if multiple mechanisms appear, split into multiple minimal quotes and separate evidence cards.
4. **Severity from firm wording only:** rely only on the firm's phrases (e.g., "materially adversely affect," "significant," "require significant investment," "may not keep pace," "disruption," "competitive disadvantage").
5. **No speculation:** if generic and not tied to internal triggers or offering-demand mechanisms, keep it only in the unclear bucket.

Evidence Card Format (All Fields Required; In Order)

For each extracted quote, output an evidence card with all fields below (each line must end with a section+page reference):

1. **Affected object (3–12 words):** identify what is at risk as named in the quote. (Section, page)
2. **One-sentence justification:** explain why this quote is obsolescence/tech-change risk and what the harm is, grounded in the quote. (Section, page)
3. **Exact quote:** "... " (Section, page)
4. **Severity signal:** choose one of {very_high, high, moderate, low} based only on wording, and cite the phrase(s) inside the quote that justify it. (Section, page)

Required Output Structure (Two Sections Only)

- **I. Internal Operating Capability Obsolescence Evidence:** include only Step 1 evidence. If none exist, output exactly:

No explicit internal operating capability obsolescence severity language found in the provided text.

- Conclude with an **Internal Severity Synthesis** (2–4 sentences). Every sentence must end with section+page references drawn from the evidence.
- **II. Customer Offering Obsolescence Evidence:** include only Step 2 evidence (after removing reassigned items); then include Step 3 under **Unclear / firm-level statements**. Conclude with an **Offering Severity Synthesis** (2–4 sentences). Every sentence must end with section+page references drawn from the evidence.

Strict Output Format (Required)

Return a valid JSON array of exactly two strings and no other text:

"<FULL Section I text (heading + evidence cards + synthesis)>",

"<FULL Section II text (heading + evidence cards + unclear subsection + synthesis)>"

Stage 2:

Input-only obsolescence prompt. **Role:** You are a financial economist.

Objective: You will be given a short summary (already extracted from a firm's 10-K). Rewrite it in simplified form to extract only statements about **input-technology obsolescence risk**—risk related to technologies used internally for operations, production, internal IT/information systems, or R&D support tools (not sold to customers).

Scope: Use only information explicitly contained in the summary. Do not use outside knowledge. Do not speculate.

Extraction Guidelines

1. Keep only severity language.

- Retain only statements describing severity / imminence / magnitude / consequences of input-technology obsolescence risk.
- Keep phrasing close to the summary; do not add interpretations.

2. Strict channel filter: input only.

- Include internal operations/production technologies, internal IT/information systems, and R&D support tools.
- Exclude anything clearly about customer-facing offerings, demand, pricing, competition, or substitution.
- If a statement is ambiguous or could be customer-facing, exclude it.

3. Preserve citations (if present).

- If the summary includes citations (e.g., section/page references), retain them.

4. Empty-case rule (verbatim output).

- If the summary contains no explicit input-technology obsolescence severity language, output exactly:

No explicit input-technology obsolescence severity language found in the provided summary.

Output Format (Required)

- Output plain text only (no JSON).
- Prefer bullet points.
- Keep phrasing close to the summary.

Output-only obsolescence prompt. *Role:* You are a financial economist.

Objective: You will be given a short summary (already extracted from a firm's 10-K). Rewrite it in simplified form to extract only statements about **output-technology obsolescence risk**—risk related to customer-facing products/services/technologies the firm sells, licenses, or delivers (including technologies embedded in offerings).

Scope: Use only information explicitly contained in the summary. Do not use outside knowledge. Do not speculate.

Extraction Guidelines

1. Keep only severity language.

- Retain only statements describing severity / imminence / magnitude / consequences of output-technology obsolescence risk.
- Keep phrasing close to the summary; do not add interpretations.

2. Strict channel filter: output only.

- Include customer-facing products/services/technologies the firm sells, licenses, or delivers (including embedded technologies).
- Exclude internal-only operational tools, internal IT/data systems, manufacturing equipment, or R&D tools used solely for operations.
- If a statement is ambiguous or could be internal-only, exclude it.

3. Firm-level / unclear channel handling.

- If the summary contains technology-change/obsolescence severity language that does not cleanly separate channels, keep it but label it as *Firm-level / unclear channel*.

4. Preserve citations (if present).

- If the summary includes citations (e.g., section/page references), retain them.

5. Empty-case rule (verbatim output).

- If the summary contains no explicit output-technology obsolescence severity language, output exactly:

No explicit output-technology obsolescence severity language found in the provided summary.

Output Format (Required)

- Output plain text only (no JSON).
- Prefer bullet points.
- Keep phrasing close to the summary.

A.5.3 Prompt to extract product information

You are an economist tasked with analyzing a firm's 10-K filing to extract and summarize all explicit mentions of **product descriptions** (including product features, uses, markets, customer needs addressed, branding, and differentiation strategies). Your goal is to identify how products are positioned, their strategic significance, and any associated challenges or opportunities.

Objective: Identify and summarize all explicit product-description content in the 10-K, focusing on product features, uses, target markets, strategic importance, and explicitly stated challenges.

Scope: Focus on **Business (Sections 1 and 1A)**.

Extraction Guidelines

1. Identify all product mentions.

- Capture every explicit mention of products, their features, uses, markets, branding, and differentiation strategies.
- Include products tied to customer engagement or strategic initiatives if explicitly described as offerings.

2. For each product, extract:

- **Name/description** as stated in the text.
- **Features/functionality** explicitly described.
- **Target market/customers** explicitly named.
- **Strategic importance/business rationale** (only if explicitly stated).
- **Challenges/risks** (only if explicitly stated).
- **Performance metrics/financial impact** (only if explicitly stated).
- **ESG factors** (only if explicitly stated).

3. Section and page references.

- Always include the section and page number for each factual point.
- If discussed across multiple sections (rare given scope), report each location.

4. Relationships between products.

- Note only explicitly stated interactions (bundles, cross-selling, complementarities).

5. Do not speculate.

- Do not infer product positioning or strategy beyond what is explicitly written.

Summary Requirements

Conclude with a brief synthesis paragraph describing the overall portfolio and any recurring themes, using only previously cited statements.

A.5.4 Sentence pairs used to compute the obsolescence risk axis

Pair 1:

- **Low Risk:** "The firm is exposed to minimal technology obsolescence risk."
- **High Risk:** "The firm is exposed to significant technology obsolescence risk."

Pair 2:

- **Low Risk:** "The company faces a low likelihood of technology becoming obsolete."
- **High Risk:** "The company faces a high likelihood of technology becoming obsolete."

Pair 3:

- **Low Risk:** "The firm encounters limited risks of technology obsolescence."
- **High Risk:** "The firm encounters severe risks of technology obsolescence."

Pair 4:

- **Low Risk:** "The company is subject to low technology obsolescence concerns."
- **High Risk:** "The company is subject to high technology obsolescence concerns."

Pair 5:

- **Low Risk:** "The firm has a minor risk of its technology becoming outdated."
- **High Risk:** "The firm has a major risk of its technology becoming outdated."

Pair 6:

- **Low Risk:** "The company experiences low exposure to risks of outdated technology."
- **High Risk:** "The company experiences high exposure to risks of outdated technology."

Pair 7:

- **Low Risk:** “The firm’s technology obsolescence risk is low.”
- **High Risk:** “The firm’s technology obsolescence risk is high.”

Pair 8:

- **Low Risk:** “The company faces negligible risk of technology obsolescence.”
- **High Risk:** “The company faces considerable risk of technology obsolescence.”

Pair 9:

- **Low Risk:** “The technology obsolescence risk for the firm is minimal.”
- **High Risk:** “The technology obsolescence risk for the firm is substantial.”

Pair 10:

- **Low Risk:** “The firm is at a low risk of technological obsolescence.”
- **High Risk:** “The firm is at a high risk of technological obsolescence.”

A.5.5 Prompt to extract specialist labor market risk

Objective. Extract only *verbatim* statements describing **current labor market conditions for high-skilled specialists** from a firm’s 10-K (not general labor market conditions for the overall workforce). “Specialists” are roles explicitly identified as skilled/technical/professional (e.g., software engineers, data scientists, R&D scientists, licensed pharmacists, board-certified/specialized nurses, certified technicians, pilots, nuclear operators, certified welders). If no qualifying specialist content appears, output exactly: *No relevant information.*

Critical requirement. Use only information explicitly stated in the 10-K. Do not infer, paraphrase, summarize, or broaden statements beyond specialist roles explicitly named or unmistakably described (e.g., “skilled engineers,” “highly specialized technicians,” “board-certified clinicians,” “PhD scientists”).

What to extract (current conditions; specialists only).

- Tightness or availability of specialist talent.
- Competition for specialist hiring, retention, or poaching.
- Wage/compensation pressure for specialist roles (e.g., sign-on/retention bonuses, equity, premium pay).
- Hiring difficulty or time-to-fill challenges for specialist positions.
- Turnover/attrition for specialists (only if explicitly framed as specialist).
- Named skill shortages in specialist areas.

Strict filtering rules.

1. The passage must explicitly reference a specialist role or use an unmistakable specialist descriptor.
2. Exclude general workforce statements unless the text explicitly ties them to specialists.
3. If a passage mixes general and specialist content, return only the specialist fragment (*verbatim*).
4. No inference, no rewording, no added context.

Required output format. Return only a numbered list of *verbatim* passages (one per item) and no additional text:

1. “[*verbatim passage 1*]”
2. “[*verbatim passage 2*]”
3. “[*verbatim passage 3*]”

A.5.6 Sentence pairs used to estimate the labor market risk

Pair 1:

- **Low Risk:** “The company reports easing wage and premium pressure for specialists (e.g., software engineers) in the current period.”
- **High Risk:** “The company reports rising wage and premium pressure for specialists (e.g., software engineers) in the current period.”

Pair 2:

- **Low Risk:** *"Hourly rates and premiums for certified technicians decreased during the reporting period."*
- **High Risk:** *"Hourly rates and premiums for certified technicians increased during the reporting period."*

Pair 3:

- **Low Risk:** *"Salary bands for R&D scientists were reduced year-over-year amid improved specialist availability."*
- **High Risk:** *"Salary bands for R&D scientists rose year-over-year due to scarce specialist supply."*

Pair 4:

- **Low Risk:** *"Time-to-fill for specialist positions (e.g., data scientists) shortened this year."*
- **High Risk:** *"Time-to-fill for specialist positions (e.g., data scientists) lengthened this year."*

Pair 5:

- **Low Risk:** *"Vacancies in specialist roles declined as a share of headcount in the current period."*
- **High Risk:** *"Vacancies in specialist roles remained elevated as a share of headcount in the current period."*

Pair 6:

- **Low Risk:** *"The firm reduced or discontinued sign-on or retention bonuses for board-certified clinicians this period."*
- **High Risk:** *"The firm implemented sign-on and retention bonuses for board-certified clinicians this period."*

Pair 7:

- **Low Risk:** *"Average contractor rates for specialized skills (e.g., nuclear operators) were lower than the prior period."*
- **High Risk:** *"Average contractor rates for specialized skills (e.g., nuclear operators) were higher than the prior period."*

Pair 8:

- **Low Risk:** *"Specialist turnover decreased and competitor poaching moderated during the year."*
- **High Risk:** *"Specialist turnover increased and competitor poaching intensified during the year."*

Pair 9:

- **Low Risk:** *"The company reports an improved pipeline of qualified licensed pharmacists for open specialist roles."*
- **High Risk:** *"The company reports difficulty securing licensed pharmacists for open specialist roles."*

Pair 10:

- **Low Risk:** *"Capacity constraints tied to specialist staffing eased in the current period."*
- **High Risk:** *"Capacity constraints tied to specialist staffing intensified in the current period."*

A.6 Example: 10-K technology summary of Microsoft in 2001

Example 1: Input-only technologies (from the summary).

1. Microsoft Research

- **Classification:** input-only (internal research; no indication of external sales).
- **Purpose:** develops innovative solutions to computer science problems; makes computers easier to use; designs software for next-generation hardware; improves the software design process.
- **Nature:** advanced, innovation-focused internal research activity.

2. Proprietary development tools and methodologies

- **Classification:** input-only (used internally for product development; not described as sold).
- **Purpose:** simplifies product portability across operating systems and microprocessors; used to create and enhance products.

- **Nature:** proprietary internal tools and processes supporting development.

3. R&D expenditures

- **Classification:** input-only (internal investment in research and development).
- **Purpose:** finances development of new products and enhancements.
- **Nature:** financial input dedicated to R&D activities.

Example 2: Customer-facing technologies (outputs; from the summary).

1. Microsoft Office

- **Classification:** product (suite of software programs sold to customers).
- **Main uses:** word processing; email; presentations; data management.
- **Noted features:** integrated applications; document-centric design; smart tags; task panes.

2. Windows operating systems (e.g., Windows XP, Windows 2000)

- **Classification:** product (operating systems sold to customers).
- **Main uses:** platform for personal computing and business applications.
- **Noted features:** enhanced reliability; security; performance for home and business users.

3. Microsoft .NET platform

- **Classification:** product (platform offered to customers).
- **Main uses:** develop/manage XML Web services; enable applications to communicate and share data over the Internet.
- **Noted features:** built on XML and Internet industry standards; provides tools for developers.

A.7 Example: Innovation summary

1. General Themes and Common Topics: The patents issued to the firm primarily focus on enhancing data management and processing systems.

- **Data reconstruction and storage efficiency:** reconstructing corrupted data from storage units without performance degradation; efficient data retrieval and redundancy.
- **Digital media processing and editing:** improving digital video delivery systems for parsing, seeking, and editing.
- **Database query optimization:** optimizing database queries for efficient record retrieval and logical evaluation.
- **Content analysis and knowledge representation:** analyzing discourse, identifying themes, and managing knowledge representation.

2. Individual Patent Details:

- **Patent 1:** corrupted data reconstruction in storage systems; real-time reconstruction using a parity block system; stripe reads and XOR-based reconstruction.
- **Patent 2:** parsing audio-visual works to create tag files; efficient seeking/editing; fast/slow forward/rewind operations; generating new video files from existing ones.
- **Patent 3:** query optimizer converting queries into optimized Boolean trees; disjoint semi-open ranges; skipping intervals based on logical conditions.
- **Patent 4:** lexicon + knowledge catalog for discourse analysis; theme parsing system; content indexing; dynamic ontology expansion.

A.8 Prompt variations

Prompt 1:

Below is a collection of patent abstracts issued to a single firm within the same year. Your task is to read the abstracts very carefully and use only the provided information to summarize them in a structured and detailed manner as follows:

1. General Themes and Common Topics:

- *Identify and summarize the overarching themes or problems addressed by multiple patents.*
- *Highlight any shared goals or innovations that appear across the abstracts.*

2. Individual Patent Details:

- *For each patent that does not fit into any of the common topics, provide a concise summary that captures its unique contributions, specific technological solutions, or methods described.*
- *Mention the targeted problem, proposed solution, and any notable technical features.*

3. Do not speculate:

- *If something is not explicitly stated, omit it. We only want the information documented in these patents.*

Here are the patent abstracts:

{patent texts}

Prompt 2:

Read the abstracts very carefully and use only the provided information to summarize them in a structured and detailed manner as follows:

1. General Themes and Common Topics:

- *Identify and summarize the overarching themes or problems addressed by multiple patents.*
- *Highlight any shared goals or innovations that appear across the abstracts.*

2. Individual Patent Details:

- *For each patent that does not fit into any of the common topics, provide a concise summary that captures its unique contributions, specific technological solutions, or methods described.*
- *Mention the targeted problem, proposed solution, and any notable technical features.*

3. Do not speculate:

- *If something is not explicitly stated, omit it. We only want the information documented in these patents.*

Here are the patent abstracts:

{patent texts}

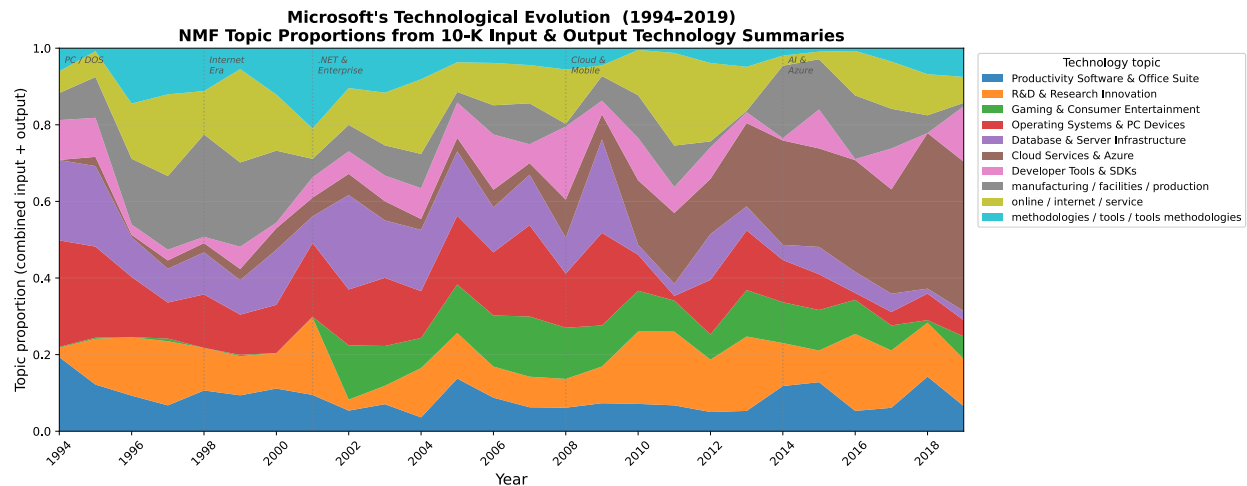


Figure B.1: This figure shows the evolution of technology topics mentioned in Microsoft’s 10-K filings from 1994 to 2019.

B Evolution of technologies over time

We use Microsoft as a case study of how the LLM-extracted process and product technology summaries shift over time, and verify that the summaries track major transitions in the firm’s 10-K narrative rather than restating boilerplate.

We construct a firm-specific corpus from Microsoft’s process- and product-technology summaries from fiscal years 1994 through 2019. Each filing year produces two structured summaries: a process-technology summary describing technologies used for internal operations and R&D, and a product-technology summary describing technologies embodied in products or services sold to customers. Because each summary is organized as a numbered list in which each entry identifies a distinct technology and provides an accompanying description, we split the summaries into item-level technology documents. We then apply a uniform cleaning procedure that removes stopwords and boilerplate phrases, and we discard items with fewer than five tokens after cleaning. This procedure yields 269 item-level documents, or 5.2 items per summary on average.

We represent each item-level document using Term Frequency–Inverse Document Frequency (TF–IDF) vectors over unigrams and bigrams. We then fit a Non-Negative Matrix Factorization (NMF) topic model, which is well-suited to short text documents. To select the number of topics, we estimate models for $K = 1, \dots, 10$ and choose the value at which the second difference in the Frobenius reconstruction error, $\|X - WH\|_F$, is maximized. This criterion selects $K = 7$ topics.

Figure B.1 shows the resulting topic evolution. Early in the sample, Microsoft’s disclosures place greater weight on operating systems and database-related technologies. Over time, the emphasis shifts toward cloud computing and related service technologies. This case study provides a concrete example of how the 10-K-based summaries track meaningful changes in a firm’s technology profile over time.

C IDE interacted with firm characteristics

Section 5.1 documents the one firm-level moderator that delivers a clean and economically meaningful interaction with composite displacement: product-market crowdedness, measured by Product

Market Similarity. A natural follow-up question is whether the IDE–profit-growth relationship is also moderated by other commonly studied firm characteristics—firm size, intangible-capital intensity, and current profitability. This appendix runs that exercise. None of these characteristics yields an interaction pattern as clean or robust as the IDE–PMS interaction in the body, so we report the results here rather than in the main text.

For each horizon $k \in \{1, \dots, 5\}$, we estimate

$$\begin{aligned} \log \Pi_{f,t+k} - \log \Pi_{f,t} = & \beta_1^k \overline{\text{IDE}}_{f,t} + \beta_2^k \text{IDE}_{f,t}^{\text{tilt}} \\ & + \beta_3^k \overline{\text{IDE}}_{f,t} \times \log(at_{f,t}) + \beta_4^k \text{IDE}_{f,t}^{\text{tilt}} \times \log(at_{f,t}) \\ & + \beta_5^k \overline{\text{IDE}}_{f,t} \times \text{IntanCap}_{f,t} + \beta_6^k \text{IDE}_{f,t}^{\text{tilt}} \times \text{IntanCap}_{f,t} \\ & + \beta_7^k \overline{\text{IDE}}_{f,t} \times \text{Profitability}_{f,t} + \beta_8^k \text{IDE}_{f,t}^{\text{tilt}} \times \text{Profitability}_{f,t} \\ & + \lambda^k \mathbf{X}_{f,t} + \delta_{s \times t}^k + \gamma_\ell^k + \epsilon_{f,t}^k, \end{aligned} \quad (28)$$

where $\log(at_{f,t})$, $\text{IntanCap}_{f,t}$, and $\text{Profitability}_{f,t}$ denote log total assets, intangible capital, and current profitability, respectively. The vector $\mathbf{X}_{f,t}$ includes these firm characteristics as controls in addition to our baseline controls. Fixed effects, clustering, and standardization follow the headline specification.

Table C.1 reports the estimates. We find two main patterns. First, the negative association between $\overline{\text{IDE}}$ and subsequent profit growth remains large and statistically significant after introducing the interaction terms. The coefficient on $\overline{\text{IDE}}$ is -0.075 at the one-year horizon and becomes more negative over longer horizons, reaching -0.242 by year five. The tilt component, IDE^{tilt} , is much smaller in magnitude and is not statistically significant at any horizon. Thus, the broad displacement component captured by $\overline{\text{IDE}}$ remains the primary source of the association with subsequent profit growth.

Second, the interaction terms are generally modest and do not point to a single dominant source of heterogeneity. Overall, these results show that the negative association between $\overline{\text{IDE}}$ and future profit growth is robust to allowing the association to vary with firm size, intangible capital, and current profitability. The interaction estimates suggest some limited heterogeneity, but the main pattern remains that the broad displacement component, $\overline{\text{IDE}}$, is consistently and negatively associated with subsequent profit growth.

D Model Appendix

This appendix develops a model linking external innovations to the focal firm’s profit growth through two channels: a product channel, operating through product-quality competition, and a process channel, operating through technology-specific factor-market pressure. The model is not intended to deliver a structural estimating equation. Instead, it identifies the economic objects that motivate our empirical IDE measures: overlap between external innovations and the focal firm’s product technologies, and overlap between external innovations and the focal firm’s process technologies.

The key feature of the process channel is that external process innovations can both raise process-specific labor demand and directly improve the productivity of innovating rivals. Thus, an external process innovation changes the focal firm’s relative position through a net cost-loading term: the focal firm faces the higher process wage while innovating rivals may also receive a process-capital productivity gain.

	Dependent variable:				
	$\log \Pi_{f,t+1} - \log \Pi_{f,t}$ (1)	$\log \Pi_{f,t+2} - \log \Pi_{f,t}$ (2)	$\log \Pi_{f,t+3} - \log \Pi_{f,t}$ (3)	$\log \Pi_{f,t+4} - \log \Pi_{f,t}$ (4)	$\log \Pi_{f,t+5} - \log \Pi_{f,t}$ (5)
$\overline{\text{IDE}} (\text{PC}_1)$	-0.075*** (0.011)	-0.131*** (0.018)	-0.167*** (0.024)	-0.208*** (0.029)	-0.242*** (0.035)
$\text{IDE}^{\text{tilt}} (\text{PC}_2)$	-0.004 (0.002)	-0.003 (0.004)	0.001 (0.005)	-0.009 (0.006)	-0.005 (0.008)
$\overline{\text{IDE}} (\text{PC}_1) \times \log(at)$	-0.002 (0.008)	-0.004 (0.013)	0.001 (0.017)	0.008 (0.021)	0.004 (0.023)
$\text{IDE}^{\text{tilt}} (\text{PC}_2) \times \log(at)$	-0.011* (0.006)	-0.011 (0.011)	-0.024* (0.013)	-0.009 (0.016)	-0.018 (0.018)
$\overline{\text{IDE}} (\text{PC}_1) \times \text{IntanCap}$	-0.005 (0.007)	-0.013 (0.012)	-0.024* (0.015)	-0.032* (0.019)	-0.032 (0.021)
$\text{IDE}^{\text{tilt}} (\text{PC}_2) \times \text{IntanCap}$	0.010* (0.006)	0.009 (0.010)	0.014 (0.012)	0.004 (0.015)	0.011 (0.016)
$\overline{\text{IDE}} (\text{PC}_1) \times \text{Profitability}$	-0.009* (0.005)	-0.003 (0.007)	0.002 (0.009)	0.016 (0.012)	0.023* (0.013)
$\text{IDE}^{\text{tilt}} (\text{PC}_2) \times \text{Profitability}$	-0.002 (0.004)	-0.002 (0.006)	-0.009 (0.008)	0.004 (0.010)	-0.006 (0.011)
Year $\times$ Sector FE	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓
Observations	48,888	44,926	41,549	38,657	36,110
R ²	0.165	0.186	0.177	0.175	0.181

TABLE C.1: This table reports the association between IDE and future profit growth after allowing the association to vary with observable firm characteristics. The dependent variable is $\log \Pi_{f,t+k} - \log \Pi_{f,t}$ for horizons $k = 1, \dots, 5$, where $\Pi_{f,t}$ denotes firm f 's profit in year t . $\overline{\text{IDE}}$ and IDE^{tilt} are the first and second principal components of process IDE and product IDE, respectively. We interact both IDE components with log total assets, intangible capital, and current profitability. We include the full set of controls: product-weighted sales growth, firm-level innovation value deflated by assets, current profitability, log current profit, log capital, intangible capital, log total assets, and log employment. IDE measures, firm-level innovation value, product-weighted sales growth, and the firm characteristics used in the interactions—log total assets, intangible capital, and current profitability—are standardized across the full sample. Fixed effects are sector-by-year and industry. Standard errors are clustered at the firm level. The data cover 1997 to 2019. All continuous variables are winsorized at the 1st and 99th percentiles within each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

D.1 Environment

There is a continuum of product markets indexed by m . In each market m , a set of active firms $\mathcal{F}_m$ competes under monopolistic competition. We work in the monopolistic-competition limit in which each firm is small relative to its market and therefore takes the market price index P_m and market expenditure E_m as given when setting its price. Firm f participates in a set of markets $\mathcal{M}_f$.

External innovations enter the focal firm's profit forecast through two channels:

- (i) **Product channel:** rivals increase product quality Q_{gm} for $g \neq f$, lowering their quality-adjusted prices and shifting expenditure shares within market m .
- (ii) **Process channel:** other firms expand process- j technology capital. Since process- j specialist labor supply is fixed in the short run, the process- j wage rises. In addition, innovating rivals may directly become more productive in process j . The focal firm's own process capital is held fixed in the local comparison.

D.2 Demand

In market m , the representative consumer aggregates quality-adjusted consumption via CES:

$$X_m = \left(\sum_{f \in \mathcal{F}_m} (Q_{fm} Y_{fm})^{\frac{\theta_m-1}{\theta_m}} \right)^{\frac{\theta_m}{\theta_m-1}}, \quad \theta_m > 1, \quad (29)$$

where Q_{fm} is product quality, Y_{fm} is physical quantity, and θ_m is the elasticity of substitution.

Define quality-adjusted quantity and quality-adjusted price:

$$\tilde{Y}_{fm} \equiv Q_{fm} Y_{fm}, \quad \tilde{p}_{fm} \equiv \frac{p_{fm}}{Q_{fm}}. \quad (30)$$

The dual CES price index is

$$P_m \equiv \left(\sum_{g \in \mathcal{F}_m} \tilde{p}_{gm}^{1-\theta_m} \right)^{\frac{1}{1-\theta_m}}. \quad (31)$$

Standard CES demand gives

$$\tilde{Y}_{fm} = X_m \left(\frac{\tilde{p}_{fm}}{P_m} \right)^{-\theta_m}. \quad (32)$$

Revenue and expenditure shares. Revenue is

$$R_{fm} \equiv p_{fm} Y_{fm} = \tilde{p}_{fm} \tilde{Y}_{fm}. \quad (33)$$

Nominal expenditure in market m is

$$E_m \equiv P_m X_m = \sum_{g \in \mathcal{F}_m} R_{gm}. \quad (34)$$

Under CES, the within-market expenditure share is

$$\sigma_{fm} \equiv \frac{R_{fm}}{E_m} = \frac{\tilde{p}_{fm}^{1-\theta_m}}{\sum_{g \in \mathcal{F}_m} \tilde{p}_{gm}^{1-\theta_m}}, \quad \sum_{g \in \mathcal{F}_m} \sigma_{gm} = 1. \quad (35)$$

Thus $R_{fm} = E_m \sigma_{fm}$.

Revenue response to relative price movements. Log-differentiating (35) gives

$$d \log \sigma_{fm} = (1 - \theta_m) (d \log \tilde{p}_{fm} - d \log P_m), \quad (36)$$

where

$$d \log P_m = \sum_{g \in \mathcal{F}_m} \sigma_{gm} d \log \tilde{p}_{gm}. \quad (37)$$

Since $R_{fm} = E_m \sigma_{fm}$,

$$d \log R_{fm} = d \log E_m + (1 - \theta_m) (d \log \tilde{p}_{fm} - d \log P_m). \quad (38)$$

Revenue growth equals market-expenditure growth plus the movement in the firm's expenditure share. The expenditure-share movement depends on the firm's quality-adjusted price relative to the market price index.

D.3 Production

CES across processes. For each firm-product pair (f, m) ,

$$Y_{fm} = \left(\sum_{j \in \mathcal{J}} y_{fm}(j)^{\frac{\psi-1}{\psi}} \right)^{\frac{\psi}{\psi-1}}, \quad \psi > 1, \quad (39)$$

where $y_{fm}(j)$ is output from process j .

Cobb-Douglas within each process. Process output is

$$y_{fm}(j) = k(f, j)^\alpha \ell_{fm}(j)^{1-\alpha}, \quad \alpha \in (0, 1), \quad (40)$$

where $k(f, j)$ is firm f 's process- j technology capital and $\ell_{fm}(j)$ is process- j labor used by firm f in market m . We interpret $k(f, j)$ as a non-rival firm-level process-technology stock that can be applied across all of the firm's product markets without depletion. The wage for process- j labor is denoted by $w(j)$.

Short-run factor-market timing. The focal firm's process-technology stocks $\{k(f, j)\}_{j \in \mathcal{J}}$ are predetermined in the short run. In the process-channel comparison below, external innovation in process j is represented as an increase in other firms' process- j technology capital, while the focal firm's own process- j capital is held fixed. Labor is specialized by process. For each process j , the short-run supply of process- j specialist labor is fixed:

$$\sum_g \sum_{m \in \mathcal{M}_g} \ell_{gm}(j) = \bar{L}_j. \quad (41)$$

Thus process- j labor markets clear independently across j . In a local process- j shock, wages in other process labor markets do not move:

$$\frac{\partial \log w(h)}{\partial z_{j,t}} = 0, \quad h \neq j. \quad (42)$$

D.4 Pricing and variable profits

Because firms are atomistic within each CES market, the perceived demand elasticity is θ_m and firms set prices as a constant markup over marginal cost:

$$p_{fm} = \frac{\theta_m}{\theta_m - 1} MC_{fm}, \quad \tilde{p}_{fm} = \frac{\theta_m}{\theta_m - 1} \frac{MC_{fm}}{Q_{fm}}. \quad (43)$$

Because the focal firm's process-technology capital is predetermined in the short run and labor is variable, the production function is homogeneous of degree $1 - \alpha$ in variable labor. Hence total variable cost satisfies

$$TC_{fm} = (1 - \alpha) MC_{fm} Y_{fm}. \quad (44)$$

Using (43),

$$MC_{fm} Y_{fm} = \frac{\theta_m - 1}{\theta_m} R_{fm}. \quad (45)$$

Therefore variable profits are

$$\Pi_{fm} = R_{fm} - TC_{fm} = \left[\alpha + \frac{1 - \alpha}{\theta_m} \right] R_{fm}. \quad (46)$$

Variable profits are a constant share of revenue in this short-run model. Because this share is constant with respect to the local movements considered below,

$$d \log \Pi_{fm} = d \log R_{fm}. \quad (47)$$

We therefore work with revenue responses throughout.

D.5 Product channel: product-quality displacement

We first derive the product-channel association. Relative to a fixed-marginal-cost benchmark, this derivation allows rival marginal costs to move endogenously through the decreasing-returns production structure introduced above.

Setup. Fix a focal firm–product pair (f, m) . Rivals' product innovations raise product qualities $\{Q_{gm}\}_{g \in \mathcal{F}_m \setminus \{f\}}$. Define

$$q_{gm} \equiv d \log Q_{gm}, \quad q_{fm} = 0. \quad (48)$$

Define the expenditure-share-weighted rival quality movement:

$$\bar{q}_m^{(-f)} \equiv \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} q_{gm}. \quad (49)$$

Since $q_{fm} = 0$, the market-level expenditure-share-weighted quality movement is

$$\bar{q}_m \equiv \sum_{g \in \mathcal{F}_m} \sigma_{gm} q_{gm} = \bar{q}_m^{(-f)}. \quad (50)$$

The quality-adjusted price is

$$\tilde{p}_{im} = \frac{p_{im}}{Q_{im}}, \quad (51)$$

so

$$d \log \tilde{p}_{im} = d \log p_{im} - q_{im}. \quad (52)$$

Physical quantity demand. CES demand implies

$$\tilde{Y}_{im} = X_m \left(\frac{\tilde{p}_{im}}{P_m} \right)^{-\theta_m}. \quad (53)$$

Since $\tilde{Y}_{im} = Q_{im} Y_{im}$,

$$Y_{im} = \frac{X_m}{Q_{im}} \left(\frac{\tilde{p}_{im}}{P_m} \right)^{-\theta_m}. \quad (54)$$

Using $E_m = P_m X_m$, log-differentiating gives

$$d \log Y_{im} = d \log E_m + (\theta_m - 1) d \log P_m - \theta_m d \log \tilde{p}_{im} - q_{im}. \quad (55)$$

Using $d \log \tilde{p}_{im} = d \log p_{im} - q_{im}$, this becomes

$$d \log Y_{im} = d \log E_m + (\theta_m - 1) d \log P_m - \theta_m d \log p_{im} + (\theta_m - 1) q_{im}. \quad (56)$$

Pricing with endogenous marginal cost. From the cost function derived below,

$$MC_{im} = \frac{1}{1 - \alpha} B_i(w, k) Y_{im}^\rho, \quad \rho \equiv \frac{\alpha}{1 - \alpha}. \quad (57)$$

In the product channel, process wages and process technology capital are held fixed, so $d \log B_i = 0$. Constant markup therefore implies

$$d \log p_{im} = d \log MC_{im} = \rho d \log Y_{im}. \quad (58)$$

Substituting (56) into (58) and solving for $d \log p_{im}$ gives

$$d \log p_{im} = \lambda_m [\rho d \log E_m + \rho(\theta_m - 1) d \log P_m + \rho(\theta_m - 1) q_{im}], \quad (59)$$

where

$$\lambda_m \equiv \frac{1}{1 + \rho \theta_m} = \frac{1 - \alpha}{1 - \alpha + \alpha \theta_m}. \quad (60)$$

Thus a quality improvement is associated with a higher physical price for the innovating firm through the output-induced marginal-cost movement. Using (52),

$$d \log \tilde{p}_{im} = \lambda_m [\rho d \log E_m + \rho(\theta_m - 1) d \log P_m - (1 + \rho) q_{im}]. \quad (61)$$

Market price-index movement. The CES price index satisfies

$$d \log P_m = \sum_{i \in \mathcal{F}_m} \sigma_{im} d \log \tilde{p}_{im}. \quad (62)$$

Using (61),

$$d \log P_m = \lambda_m [\rho d \log E_m + \rho(\theta_m - 1) d \log P_m - (1 + \rho) \bar{q}_m]. \quad (63)$$

Solving for $d \log P_m$ gives

$$d \log P_m = \alpha d \log E_m - \bar{q}_m. \quad (64)$$

Since $\bar{q}_m = \bar{q}_m^{(-f)}$,

$$d \log P_m = \alpha d \log E_m - \bar{q}_m^{(-f)}. \quad (65)$$

Focal firm's relative quality-adjusted price. For the focal firm, $q_{fm} = 0$. Therefore,

$$d \log \tilde{p}_{fm} = \lambda_m [\rho d \log E_m + \rho(\theta_m - 1) d \log P_m]. \quad (66)$$

Combining this expression with (65) gives

$$d \log \tilde{p}_{fm} - d \log P_m = \eta_m \bar{q}_m^{(-f)}, \quad (67)$$

where

$$\eta_m \equiv \lambda_m(1 + \rho) = \frac{1}{1 + \alpha(\theta_m - 1)} \in (0, 1]. \quad (68)$$

The focal firm's relative quality-adjusted price rises when rivals improve quality. The magnitude is dampened by η_m . When $\alpha = 0$, marginal cost does not rise with output and $\eta_m = 1$.

Fixed-expenditure benchmark. Using (38) and imposing $d \log E_m = 0$,

$$d \log R_{fm} = (1 - \theta_m) (d \log \tilde{p}_{fm} - d \log P_m). \quad (69)$$

Substituting (67),

$$d \log R_{fm} = -(\theta_m - 1)\eta_m \bar{q}_m^{(-f)}. \quad (70)$$

Because $d \log \Pi_{fm} = d \log R_{fm}$ for variable profits,

$$d \log \Pi_{fm}^{\text{prod}} = -(\theta_m - 1)\eta_m \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} d \log Q_{gm}. \quad (71)$$

General case with expenditure reallocation. Allow market expenditure to move with the market price index. Under a local upper-tier CES benchmark, write

$$d \log E_m = -\Lambda_m d \log P_m, \quad \Lambda_m \equiv (\nu - 1)(1 - \vartheta_m) \geq 0. \quad (72)$$

Here ν is the elasticity of substitution across markets in the upper tier and ϑ_m is market m 's expenditure share. We impose the standard nested ordering $1 < \nu < \theta_m$: varieties within a market substitute more readily than markets substitute for one another, and expenditure responds normally to prices, flowing toward a market whose price index falls. The ordering delivers $0 \leq \Lambda_m < \theta_m - 1$, since $\vartheta_m \in (0, 1)$ implies $\Lambda_m < \nu - 1$. This expression applies when market m is the only market whose price index changes locally. When multiple markets move together, the corresponding upper-tier CES expression involves $d \log P_m - \sum_n \vartheta_n d \log P_n$. Substituting (72) into (65) gives

$$d \log P_m = -\frac{1}{1 + \alpha \Lambda_m} \bar{q}_m^{(-f)}. \quad (73)$$

Therefore,

$$d \log E_m = \frac{\Lambda_m}{1 + \alpha \Lambda_m} \bar{q}_m^{(-f)}. \quad (74)$$

The focal firm's relative quality-adjusted price remains

$$d \log \tilde{p}_{fm} - d \log P_m = \eta_m \bar{q}_m^{(-f)}. \quad (75)$$

Hence the product-channel profit-growth prediction is

$$d \log \Pi_{fm}^{\text{prod}} = - \left[(\theta_m - 1)\eta_m - \frac{\Lambda_m}{1 + \alpha \Lambda_m} \right] \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} d \log Q_{gm}. \quad (76)$$

The bracket signs the channel. Since $\eta_m = 1/(1 + \alpha(\theta_m - 1))$, it equals $h(\theta_m - 1) - h(\Lambda_m)$ with $h(x) \equiv x/(1 + \alpha x)$ increasing, so it is positive—rival quality gains reduce focal profit—if and only if $\Lambda_m < \theta_m - 1$, which the nested ordering $\nu < \theta_m$ guarantees. The two terms are the competing forces

a rival quality improvement sets off: business stealing within the market, of strength $(\theta_m - 1)\eta_m$, against expenditure reallocation into the market as its price index falls, of strength $\Lambda_m/(1 + \alpha\Lambda_m)$. Displacement requires the first to dominate; were markets to substitute more readily than varieties within them, the focal firm would gain from a rival's improvement.

Firm-level aggregation. Total firm profits are $\Pi_f = \sum_{m \in \mathcal{M}_f} \Pi_{fm}$. Define profit weights

$$\pi_{fm} \equiv \frac{\Pi_{fm}}{\Pi_f}, \quad \sum_{m \in \mathcal{M}_f} \pi_{fm} = 1. \quad (77)$$

Under fixed market expenditure,

$$d \log \Pi_f^{\text{prod}} = - \sum_{m \in \mathcal{M}_f} \pi_{fm} (\theta_m - 1) \eta_m \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} d \log Q_{gm}. \quad (78)$$

More generally,

$$d \log \Pi_f^{\text{prod}} = - \sum_{m \in \mathcal{M}_f} \pi_{fm} \left[(\theta_m - 1) \eta_m - \frac{\Lambda_m}{1 + \alpha \Lambda_m} \right] \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} d \log Q_{gm}. \quad (79)$$

D.6 Process channel: factor-market displacement

We next derive the process-channel association. Product qualities $\{Q_{gm}\}$ are held fixed throughout this subsection. External innovation in process j is represented as an increase in other firms' effective process- j technology capital, holding the focal firm's own process- j capital fixed. Since process- j specialist labor supply is fixed in the short run, the process- j wage rises. When some of the innovating firms also compete with the focal firm in market m , their own process-capital gains enter their cost responses and must be carried through.

D.6.1 Cost minimization and process exposure shares

For a given output target Y_{fm} , firm f chooses labor allocations $\{\ell_{fm}(j)\}_{j \in \mathcal{J}}$ to minimize total variable cost

$$\sum_{j \in \mathcal{J}} w(j) \ell_{fm}(j) \quad (80)$$

subject to (39)–(40). Substituting (40) into (39) gives

$$Y_{fm} = \left(\sum_{j \in \mathcal{J}} k(f, j)^{\frac{\alpha(\psi-1)}{\psi}} \ell_{fm}(j)^{\frac{(1-\alpha)(\psi-1)}{\psi}} \right)^{\frac{\psi}{\psi-1}}. \quad (81)$$

For each process, the labor requirement needed to produce $y_{fm}(j)$ is

$$\ell_{fm}(j) = \left(\frac{y_{fm}(j)}{k(f, j)^\alpha} \right)^{\frac{1}{1-\alpha}}. \quad (82)$$

Thus the cost of obtaining $y_{fm}(j)$ units of process- j output is

$$w(j)k(f, j)^{-\frac{\alpha}{1-\alpha}} y_{fm}(j)^{\frac{1}{1-\alpha}}. \quad (83)$$

Let

$$\tilde{\psi} \equiv \frac{\psi}{1 + \alpha(\psi - 1)}. \quad (84)$$

The cost-minimization first-order conditions imply

$$y_{fm}(j) \propto \left(\frac{w(j)}{k(f, j)^{\alpha/(1-\alpha)}} \right)^{-\tilde{\psi}(1-\alpha)}. \quad (85)$$

Define the process exposure share as the process- j variable-cost share:

$$s_{f,j} \equiv \frac{w(j)\ell_{fm}(j)}{\sum_{h \in \mathcal{J}} w(h)\ell_{fm}(h)}. \quad (86)$$

Because the production function is homothetic and all products of the same firm use the same process aggregator and the same firm-level process capital vector, this share does not depend on output scale or on the product market m . We therefore write it as $s_{f,j}$. Using (85), the exposure share can be written as

$$s_{f,j} = \frac{\left[k(f, j)^{\alpha/(1-\alpha)} / w(j) \right]^{\tilde{\psi}-1}}{\sum_{h \in \mathcal{J}} \left[k(f, h)^{\alpha/(1-\alpha)} / w(h) \right]^{\tilde{\psi}-1}}. \quad (87)$$

The exposure share is higher when the firm has more technology capital in process j and lower when process- j labor is more expensive.

Variable cost function and marginal cost. Because the production function is homogeneous of degree $1 - \alpha$ in variable labor, the minimized variable cost function takes the form

$$TC_{fm} = B_f(w, k) Y_{fm}^{1/(1-\alpha)}, \quad (88)$$

where $B_f(w, k)$ is a cost index that depends on the vector of process wages $w \equiv \{w(j)\}_{j \in \mathcal{J}}$ and the firm's process capital vector $k \equiv \{k(f, j)\}_{j \in \mathcal{J}}$ but not on output scale. Marginal cost is therefore

$$MC_{fm} = \frac{1}{1-\alpha} B_f(w, k) Y_{fm}^{\rho}, \quad \rho \equiv \frac{\alpha}{1-\alpha}. \quad (89)$$

By the envelope theorem, the proportional movement in the cost index induced by small wage and process-capital movements, holding output fixed, is

$$d \log B_f = \sum_{j \in \mathcal{J}} s_{f,j} d \log w(j) - \rho \sum_{j \in \mathcal{J}} s_{f,j} d \log k(f, j). \quad (90)$$

The first term is the wage-cost component. The second term is the process-capital productivity component. For the focal firm in the local comparison below, $d \log k(f, j) = 0$.

The full marginal-cost movement combines the direct cost-index component with output feedback:

$$d \log MC_{fm} = d \log B_f + \rho d \log Y_{fm}. \quad (91)$$

D.6.2 External process capital and the process-specific wage

Let $\chi_{gm,j,t}$ denote the shadow value of one unit of process- j output for firm g producing in market m . Holding these shadow values fixed in the local wage-clearing step, firm g 's demand for process- j labor solves

$$\max_{\ell_{gm}(j)} \chi_{gm,j,t} k(g,j)^\alpha \ell_{gm}(j)^{1-\alpha} - w(j) \ell_{gm}(j). \quad (92)$$

The first-order condition is

$$w(j) = (1 - \alpha) \chi_{gm,j,t} k(g,j)^\alpha \ell_{gm}(j)^{-\alpha}. \quad (93)$$

Solving for process- j labor demand gives

$$\ell_{gm}(j) = \left(\frac{(1 - \alpha) \chi_{gm,j,t}}{w(j)} \right)^{1/\alpha} k(g,j). \quad (94)$$

Aggregating across all firms and product markets, process- j labor-market clearing requires

$$\bar{L}_j = \sum_g \sum_{m \in \mathcal{M}_g} \left(\frac{(1 - \alpha) \chi_{gm,j,t}}{w(j)} \right)^{1/\alpha} k(g,j). \quad (95)$$

Define the effective process- j technology capital stock as

$$K_{j,t}^{\text{eff}} \equiv \sum_g \sum_{m \in \mathcal{M}_g} \chi_{gm,j,t}^{1/\alpha} k(g,j). \quad (96)$$

Then (95) implies

$$w(j) = (1 - \alpha) \left(\frac{K_{j,t}^{\text{eff}}}{\bar{L}_j} \right)^\alpha. \quad (97)$$

Since $\bar{L}_j$ is fixed in the short run,

$$d \log w(j) = \alpha d \log K_{j,t}^{\text{eff}}. \quad (98)$$

Decompose effective process- j capital into the focal firm's component and the other-firm component:

$$K_{j,t}^{\text{eff}} = K_{j,t}^{(f),\text{eff}} + K_{j,t}^{(-f),\text{eff}}, \quad (99)$$

where

$$K_{j,t}^{(f),\text{eff}} \equiv \sum_{m \in \mathcal{M}_f} \chi_{fm,j,t}^{1/\alpha} k(f,j), \quad (100)$$

and

$$K_{j,t}^{(-f),\text{eff}} \equiv \sum_{g \neq f} \sum_{m \in \mathcal{M}_g} \chi_{gm,j,t}^{1/\alpha} k(g,j). \quad (101)$$

Let the external process- j IDE shock be the small log movement in other firms' effective process- j capital:

$$z_{j,t} \equiv d \log K_{j,t}^{(-f),\text{eff}}. \quad (102)$$

The focal firm's own process- j capital is held fixed in this local comparison:

$$d \log K_{j,t}^{(f),\text{eff}} = 0. \quad (103)$$

Using (98), the local wage association with external process- j capital is

$$\frac{\partial \log w(j)}{\partial z_{j,t}} = \alpha \frac{K_{j,t}^{(-f),\text{eff}}}{K_{j,t}^{\text{eff}}} \equiv \kappa_{j,t} > 0. \quad (104)$$

For $h \neq j$, process- h labor is segmented from process- j labor, so its wage does not move in this local process- j comparison:

$$\frac{\partial \log w(h)}{\partial z_{j,t}} = 0, \quad h \neq j. \quad (105)$$

Combining (104) and (105),

$$\frac{\partial \log w(h)}{\partial z_{j,t}} = \kappa_{j,t} \mathbf{1}\{h = j\}, \quad \kappa_{j,t} \equiv \alpha \frac{K_{j,t}^{(-f),\text{eff}}}{K_{j,t}^{\text{eff}}} > 0. \quad (106)$$

Thus the local wage association is governed by α and by the other-firm share of effective process- j capital. This derivation holds process shadow values $\chi_{gm,j,t}$ fixed. If downstream prices, quantities, and shadow values also move contemporaneously, then $\kappa_{j,t}$ should be interpreted as a reduced-form local wage elasticity linking external process- j IDE shocks to the process- j wage.

D.6.3 Net process-cost loadings

Under the process-channel comparison, the focal firm's own process capital is fixed:

$$\frac{\partial \log k(f,j)}{\partial z_{j,t}} = 0. \quad (107)$$

However, some rival firms may directly receive the process- j technology-capital improvement. Define the firm-specific process-capital response:

$$e_{g,j,t} \equiv \frac{\partial \log k(g,j)}{\partial z_{j,t}}. \quad (108)$$

For the focal firm, $e_{f,j,t} = 0$. For an innovating rival, $e_{g,j,t}$ may be positive.

Using (90) and (106), the direct cost-index response of firm g to the external process- j IDE shock is

$$\frac{\partial \log B_g}{\partial z_{j,t}} = s_{g,j} \kappa_{j,t} - \rho s_{g,j} e_{g,j,t}. \quad (109)$$

Define the net direct process-cost loading:

$$b_{g,j,t} \equiv s_{g,j} (\kappa_{j,t} - \rho e_{g,j,t}), \quad \rho \equiv \frac{\alpha}{1 - \alpha}. \quad (110)$$

For the focal firm,

$$b_{f,j,t} = \kappa_{j,t} s_{f,j}. \quad (111)$$

For an innovating rival, $b_{g,j,t}$ can be positive, zero, or negative. It is positive when the wage-pressure component dominates the process-capital productivity component, and negative when the process-capital productivity component dominates the wage-pressure component.

Define the expenditure-weighted market average net process-cost loading:

$$\bar{b}_{m,j,t} \equiv \sum_{g \in \mathcal{F}_m} \sigma_{gm} b_{g,j,t}. \quad (112)$$

D.6.4 Equilibrium pricing response

Because product qualities are held fixed in the process channel,

$$d \log \tilde{p}_{gm} = d \log p_{gm}. \quad (113)$$

Constant markup implies

$$d \log p_{gm} = d \log MC_{gm}. \quad (114)$$

Using (91),

$$d \log p_{gm} = d \log B_g + \rho d \log Y_{gm}. \quad (115)$$

The net cost loading (110) gives

$$d \log B_g = b_{g,j,t} z_{j,t}. \quad (116)$$

CES demand for physical quantity gives

$$d \log Y_{gm} = d \log E_m + (\theta_m - 1) d \log P_m - \theta_m d \log p_{gm}. \quad (117)$$

Substituting (117) into (115) and solving for $d \log p_{gm}$ gives

$$d \log p_{gm} = \lambda_m [b_{g,j,t} z_{j,t} + \rho d \log E_m + \rho(\theta_m - 1) d \log P_m], \quad (118)$$

where

$$\lambda_m \equiv \frac{1}{1 + \rho \theta_m} = \frac{1 - \alpha}{1 - \alpha + \alpha \theta_m} \in (0, 1). \quad (119)$$

The factor λ_m summarizes the output-feedback dampening associated with decreasing returns in variable labor.

D.6.5 Market price-index response

Taking the expenditure-weighted average of (118) across firms in market m gives

$$d \log P_m = \lambda_m [\bar{b}_{m,j,t} z_{j,t} + \rho d \log E_m + \rho(\theta_m - 1) d \log P_m]. \quad (120)$$

Solving for $d \log P_m$,

$$d \log P_m = (1 - \alpha) [\bar{b}_{m,j,t} z_{j,t} + \rho d \log E_m], \quad (121)$$

where we have used

$$\frac{\lambda_m}{1 - \lambda_m \rho(\theta_m - 1)} = \frac{1}{1 + \rho} = 1 - \alpha. \quad (122)$$

Relative price response. Subtracting the market average from (118), the expenditure and price-index terms cancel:

$$d \log p_{fm} - d \log P_m = \lambda_m (b_{f,j,t} - \bar{b}_{m,j,t}) z_{j,t}. \quad (123)$$

The relative-price movement depends on the focal firm's net cost loading relative to the market average.

D.6.6 Product-level profit-growth prediction

Using $d \log \Pi_{fm} = d \log R_{fm}$ from (47), the revenue response (38), and the relative-price expression (123), we obtain

$$d \log \Pi_{fm}^{\text{proc}} = d \log E_m - (\theta_m - 1) \lambda_m (b_{f,j,t} - \bar{b}_{m,j,t}) z_{j,t}. \quad (124)$$

Equivalently,

$$\frac{\partial \log \Pi_{fm}^{\text{proc}}}{\partial z_{j,t}} = \frac{\partial \log E_m}{\partial z_{j,t}} - (\theta_m - 1) \lambda_m (b_{f,j,t} - \bar{b}_{m,j,t}). \quad (125)$$

Fixed-expenditure benchmark. Under the fixed-expenditure benchmark,

$$\frac{\partial \log E_m}{\partial z_{j,t}} = 0. \quad (126)$$

The process-channel prediction becomes

$$\frac{\partial \log \Pi_{fm}^{\text{proc}}}{\partial z_{j,t}} = -(\theta_m - 1) \lambda_m (b_{f,j,t} - \bar{b}_{m,j,t}). \quad (127)$$

Thus the process-channel prediction is negative when the focal firm's net process-cost loading exceeds the market average:

$$b_{f,j,t} > \bar{b}_{m,j,t}. \quad (128)$$

Because $b_{f,j,t} = \kappa_{j,t} s_{f,j}$ while innovating rivals may have

$$b_{g,j,t} = s_{g,j} (\kappa_{j,t} - \rho e_{g,j,t}), \quad (129)$$

external process innovations can worsen the focal firm's relative cost position both by raising the process- j wage and by improving rivals' process productivity.

Allowing market expenditure to move. More generally, process-side cost pressure can move total expenditure in market m through the market price index. Under a local upper-tier CES benchmark, write

$$d \log E_m = -\Lambda_m d \log P_m, \quad \Lambda_m \equiv (\nu - 1)(1 - \vartheta_m) \geq 0. \quad (130)$$

Here ν is the elasticity of substitution across markets in the upper tier and ϑ_m is market m 's expenditure share. We impose the standard nested ordering $1 < \nu < \theta_m$: varieties within a market substitute more readily than markets substitute for one another, and expenditure responds normally to prices, flowing toward a market whose price index falls. The ordering delivers $0 \leq \Lambda_m < \theta_m - 1$, since $\vartheta_m \in (0, 1)$ implies $\Lambda_m < \nu - 1$. This expression applies when market m is the only market whose price index changes locally. When multiple markets move together, the corresponding upper-tier CES expression involves $d \log P_m - \sum_n \vartheta_n d \log P_n$. Substituting $d \log E_m = -\Lambda_m d \log P_m$

into (121) and solving gives

$$d \log P_m = \frac{1 - \alpha}{1 + \alpha \Lambda_m} \bar{b}_{m,j,t} z_{j,t}. \quad (131)$$

Hence

$$d \log E_m = -\frac{(1 - \alpha) \Lambda_m}{1 + \alpha \Lambda_m} \bar{b}_{m,j,t} z_{j,t}. \quad (132)$$

Equivalently,

$$\frac{\partial \log E_m}{\partial z_{j,t}} = -\frac{(1 - \alpha) \Lambda_m}{1 + \alpha \Lambda_m} \bar{b}_{m,j,t}. \quad (133)$$

Substituting into (125) gives the general process-channel prediction:

$$\frac{\partial \log \Pi_{fm}^{\text{proc}}}{\partial z_{j,t}} = \underbrace{-\frac{(1 - \alpha) \Lambda_m}{1 + \alpha \Lambda_m} \bar{b}_{m,j,t}}_{\text{market-expenditure component}} + \underbrace{-(\theta_m - 1) \lambda_m (b_{f,j,t} - \bar{b}_{m,j,t})}_{\text{within-market relative-cost component}}. \quad (134)$$

Interpretation. Equation (134) shows that the process-channel prediction depends on the focal firm's net cost loading, $b_{f,j,t}$; the market-average net cost loading, $\bar{b}_{m,j,t}$; the local wage elasticity of process- j specialist labor, $\kappa_{j,t}$; rival process-capital responses, $e_{g,j,t}$; and the sensitivity of market expenditure to the market price index, Λ_m . The within-market component depends on the focal firm's net cost loading relative to the market average, dampened by λ_m . The market-expenditure component depends on the average net cost loading of firms in the market.

D.6.7 Firm-level aggregation

Aggregation across products. Using profit weights $\pi_{fm} \equiv \Pi_{fm} / \Pi_f$, the firm-level elasticity to the process- j IDE shock is

$$\frac{\partial \log \Pi_f^{\text{proc}}}{\partial z_{j,t}} = \sum_{m \in \mathcal{M}_f} \pi_{fm} \frac{\partial \log \Pi_{fm}^{\text{proc}}}{\partial z_{j,t}}. \quad (135)$$

Aggregation across processes. The total process-channel profit-growth prediction for firm f is

$$d \log \Pi_f^{\text{proc}} = \sum_{j \in \mathcal{J}} \left[\sum_{m \in \mathcal{M}_f} \pi_{fm} \frac{\partial \log \Pi_{fm}^{\text{proc}}}{\partial z_{j,t}} \right] z_{j,t}. \quad (136)$$

Equivalently, substituting (134),

$$d \log \Pi_f^{\text{proc}} = - \sum_{j \in \mathcal{J}} z_{j,t} \left[\sum_{m \in \mathcal{M}_f} \pi_{fm} \left\{ \frac{(1 - \alpha) \Lambda_m}{1 + \alpha \Lambda_m} \bar{b}_{m,j,t} + (\theta_m - 1) \lambda_m (b_{f,j,t} - \bar{b}_{m,j,t}) \right\} \right]. \quad (137)$$

This expression makes explicit the double aggregation over products and processes.

D.7 Total displacement association and mapping to empirical IDE

The total profit-growth prediction is the sum of the product and process components:

$$d \log \Pi_f = d \log \Pi_f^{\text{prod}} + d \log \Pi_f^{\text{proc}}. \quad (138)$$

The product channel depends on the overlap between external innovations and the focal firm's product technologies. In the fixed-expenditure benchmark, the relevant expression is

$$\sum_{m \in \mathcal{M}_f} \pi_{fm} (\theta_m - 1) \eta_m \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} d \log Q_{gm}, \quad (139)$$

where η_m captures the dampening induced by endogenous marginal-cost movements. More generally, the product-channel expression also includes the market-expenditure component governed by Λ_m .

The process channel depends on the overlap between external innovations and the focal firm's process technologies. With net cost loadings, the relevant process-side object is not only the focal firm's exposure $s_{f,j}$ but also the market-average net loading:

$$\bar{b}_{m,j,t} = \sum_{g \in \mathcal{F}_m} \sigma_{gm} s_{g,j} \left(\kappa_{j,t} - \frac{\alpha}{1 - \alpha} e_{g,j,t} \right). \quad (140)$$

The focal firm's own net loading is

$$b_{f,j,t} = \kappa_{j,t} s_{f,j}, \quad (141)$$

because the focal firm's own process capital is held fixed. Innovating rivals can have lower net loadings, or even negative net loadings, when their process-capital productivity gains dominate the common wage-pressure component.

Our empirical construction measures the focal-firm exposure component directly. For each firm f and year t , we compute

$$\text{IDE}_{f,t}^{\text{process}} = \sum_{g \neq f} \cos(\text{Innov}_{g,t}, \text{Tech}_{f,t}^{\text{process}}) A_{g,t}, \quad (142)$$

and

$$\text{IDE}_{f,t}^{\text{product}} = \sum_{g \neq f} \cos(\text{Innov}_{g,t}, \text{Tech}_{f,t}^{\text{product}}) A_{g,t}, \quad (143)$$

where g indexes external innovating firms and $A_{g,t}$ is the economic value of firm g 's innovations, scaled by firm size. These measures aggregate valuable external innovations by their technological proximity to the focal firm's process and product stacks.

These measures are empirical proxies for focal-firm exposure to valuable external innovations, not structural sufficient statistics for the full profit-growth mapping. The model motivates why product and process technology overlap should forecast profit growth, but the regression design absorbs broad industry and sector-year variation, so the relevant variation comes from differences in exposure among firms operating in similar product and time environments. The additional equilibrium objects— η_m , λ_m , $\bar{b}_{m,j,t}$, $\kappa_{j,t}$, $e_{g,j,t}$, and Λ_m —are not separately observed in the empirical measures. The regressions should therefore be interpreted as reduced-form associations between focal-firm exposure to valuable external innovations and subsequent profit growth, with these unobserved equilibrium objects entering the coefficient as heterogeneous pass-through and market-

response components.

E Additional analyses

In this section, we show two additional analyses. First, we show that our composite IDE measure is also significantly associated with the focal firm’s profitability, defined as $\frac{sale_{f,t} - cogs_{f,t}}{at_{f,t}}$. Then, we provide benchmark magnitudes for our main result using R&D expenditures and capital expenditures.

E.1 IDE and profitability growth

We replicate our baseline regression where we use profitability as the dependent variable instead of profit. In particular, we study the association between IDE and the log growth in profitability under the same controls, fixed effects, winsorization, and clustering as equation 19.

Firm profitability growth $\log(\Pi/A)_{f,t+h} - \log(\Pi/A)_{f,t}$	Horizon (years)				
	(1)	(2)	(3)	(4)	(5)
IDE (PC ₁)	-0.029*** (0.010)	-0.053*** (0.014)	-0.051*** (0.017)	-0.054*** (0.020)	-0.044** (0.022)
IDE ^{tilt} (PC ₂)	-0.000 (0.002)	0.000 (0.003)	0.004 (0.003)	0.001 (0.004)	0.002 (0.004)
Firm innovation	0.008*** (0.002)	0.013*** (0.004)	0.014*** (0.005)	0.012** (0.005)	0.014** (0.006)
Observations	50,025	45,969	42,521	39,567	36,963
R ²	0.127	0.175	0.193	0.209	0.227
Fixed Effects:					
Year × Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓

TABLE E.1: IDE and firm profitability. This table shows the association between our IDE measure and profitability ($\frac{sale_{f,t} - cogs_{f,t}}{at_{f,t}}$) growth over the next five years (one to five years ahead). $\overline{IDE}$ (PC₁) is the composite displacement and IDE^{tilt} (PC₂) is the asymmetry component (the first and second principal components, respectively, of process and product IDE). We include the baseline controls: current profitability, defined as operating profits (Compustat sale minus cogs) divided by book assets (at); log of current profit; current capital stock (Compustat ppeg, in logs); current employment (Compustat emp, in logs); and product-weighted sales growth—the product-similarity-weighted average of contemporaneous sales growth across all other firms, where the weights are pairwise product similarities computed from our own 10-K embeddings following [Hoberg and Phillips \(2016\)](#). IDE measures, firm-level innovation value, and product-weighted sales growth are standardized across the entire sample. Fixed effects are sector-by-year and industry. Standard errors are clustered at the firm level. The sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

As shown in Table E.1, our composite IDE is significantly negatively associated with the profitability growth of the focal firm. In particular, one standard deviation increase in composite IDE is associated with a 2.9 log point decrease in profitability growth by the next year and between 4.4 and 5.4 log points at horizons of two to five years. This shows that firms that are exposed to higher innovation displacement from other firms tend to grow more slowly in profitability.

E.2 Benchmarking the association between IDE and profit growth

To better interpret the size of the association between IDE and profit growth, we compare the coefficients against two variables that represent firms' own growth potential: logged R&D expenditure and logged capital expenditure. We replicate our baseline regression, adding these two controls and removing the firm's own innovation value — since it may absorb some information from R&D expenditures. As shown in Table E.2, the positive association between capital expenditures and profit growth is similar in magnitude to our composite IDE measure, and both are systematically larger than the association between R&D expenditures and profit growth. This shows that the association between composite IDE and the focal firm's profit growth is economically large. Both benchmarking comparisons restrict the sample to firm-years with strictly positive R&D and capital expenditure, since both enter in logs; N at one year falls from 50,028 to 25,838 here and to 24,791 in Table E.3. Composite IDE is -0.105 to -0.285 on the first subsample and -0.104 to -0.296 on the second, against -0.106 to -0.290 in the headline specification, so the restriction costs half the sample and leaves the coefficient essentially unchanged.

Firm profit growth $\log \Pi_{f,t+h} - \log \Pi_{f,t}$	Horizon (years)				
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}} \text{ (PC}_1\text{)}$	-0.105^{***} (0.015)	-0.175^{***} (0.025)	-0.195^{***} (0.033)	-0.229^{***} (0.040)	-0.285^{***} (0.048)
$\text{IDE}^{\text{tilt}} \text{ (PC}_2\text{)}$	-0.001 (0.003)	0.001 (0.004)	0.009 (0.006)	0.006 (0.007)	0.010 (0.009)
log R&D exp. (std.)	0.061^{***} (0.007)	0.108^{***} (0.011)	0.135^{***} (0.015)	0.156^{***} (0.019)	0.177^{***} (0.022)
log Capital exp. (std.)	0.107^{***} (0.011)	0.190^{***} (0.017)	0.262^{***} (0.023)	0.290^{***} (0.028)	0.316^{***} (0.033)
Year $\times$ Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓
Observations	25,838	23,823	22,091	20,551	19,186
R^2	0.134	0.164	0.168	0.168	0.177

TABLE E.2: Benchmarking the economic magnitude of IDE against own R&D and capital investment. Each column regresses profit growth on $\overline{\text{IDE}}$ (PC₁), IDE^{tilt} (PC₂), $\log(\text{xrd})$ and $\log(\text{capx})$, and the baseline controls (product-weighted sales growth, log capital, log employment, log profit, profitability). IDE measures, log R&D expenditures, capital expenditures, and product-similarity-weighted sales growth are standardized across the whole sample. Fixed effects are year $\times$ sector and industry. Standard errors are clustered at the firm level. Because both benchmarks enter in logs, the sample is restricted to firm-years with strictly positive R&D and capital expenditure, which reduces the number of observations from 50,028 to 25,838 at the one-year horizon; the sample period covers 1997 to 2019. All firm-level variables are winsorized at the 1/99% level in each year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A natural further benchmark is the R&D spending of the focal firm's industry peers. It is the comparison closest in spirit to IDE, since both are built from what *other* firms do rather than from the focal firm's own activity. We measure peer R&D as the average R&D expenditure of the other Compustat firms in the focal firm's 3-digit SIC industry and year, requiring at least two reporting peers, and enter it as $\log(\cdot)$ — the same transform, on the same strictly-positive requirement,

that own R&D and own capital expenditure receive — winsorized by year and standardized. Table E.3 reports the comparison on the sample where all four measures are available, with every regressor standardized over that sample so the coefficients are per standard deviation and directly comparable.

One specification choice matters. Peer R&D varies only across industry-years, so the 3-digit SIC fixed effect used elsewhere absorbs almost all of its variation and would leave the comparison to be made against a regressor that has been partialled away. Panel A therefore retains sector and sector-by-year effects but drops the industry absorber; Panel B restores it, to show exactly what it does.

Without the industry absorber, peer R&D enters negatively and significantly from one to four years, and marginally at five: firms whose industry peers spend more on R&D do go on to grow profits more slowly, which is the sign displacement predicts. Its magnitude, however, is an order of magnitude below composite IDE on the same sample, and including it leaves the IDE coefficient unchanged. Restoring the industry absorber in Panel B leaves the peer-R&D point estimates similar but removes their statistical significance, as expected once the variation they rest on has been absorbed.

The ranking is the same in both exercises and in Table E.2: composite IDE and capital expenditure carry the two largest associations and are similar in size, composite IDE is one and a half to two times own R&D, and peer R&D is an order of magnitude smaller. Displacement exposure is therefore not a restatement of how much the firm's industry spends on research. Spending is an input; IDE measures the realized and valuable innovation of others, weighted by how close it sits to what the focal firm actually does.

E.3 Fully interacted moderator specification

Equation (22) interacts product-market similarity with the two IDE components but not with the control vector. Panel A below reproduces that specification; Panel B additionally interacts every control with the moderator, so that no covariate is restricted to a homogeneous association across the similarity distribution. The two panels are estimated on the same sample and the coefficients are almost identical, so the interaction is not an artifact of the homogeneity restriction.

Firm profit growth	Horizon (years)				
$\log \Pi_{f,t+h} - \log \Pi_{f,t}$	(1)	(2)	(3)	(4)	(5)
<i>Panel A: sector and sector-by-year effects, no industry absorber</i>					
$\overline{\text{IDE}} (\text{PC}_1)$	−0.104***	−0.170***	−0.191***	−0.236***	−0.296***
Peer R&D (std.)	−0.009***	−0.019***	−0.020**	−0.023**	−0.023*
log R&D exp. (std.)	0.048***	0.088***	0.108***	0.124***	0.144***
log Capital exp. (std.)	0.099***	0.174***	0.243***	0.265***	0.289***
Observations	24,791	22,849	21,182	19,705	18,396
<i>Panel B: industry absorber restored</i>					
$\overline{\text{IDE}} (\text{PC}_1)$	−0.104***	−0.171***	−0.189***	−0.224***	−0.283***
Peer R&D (std.)	−0.009	−0.015	−0.021	−0.020	−0.015
log R&D exp. (std.)	0.052***	0.093***	0.117***	0.135***	0.154***
log Capital exp. (std.)	0.100***	0.176***	0.244***	0.269***	0.292***
Observations	24,783	22,843	21,174	19,696	18,386
Year × Sector	✓	✓	✓	✓	✓
Industry (3-digit SIC)			Panel B only		
Baseline controls	✓	✓	✓	✓	✓

TABLE E.3: Benchmarking composite IDE against peer R&D, own R&D and capital investment. Peer R&D is the average R&D expenditure of the other Compustat firms in the focal firm's 3-digit SIC industry and year, requiring at least two reporting peers. The peer average is formed first and then logged, so peer R&D receives the same transform as own R&D and own capital expenditure. All four regressors are winsorized at 1/99 within each year and standardized over the common sample on which each is available, so coefficients are per standard deviation. IDE^{tilt} and the baseline controls are included but not shown. Panel A absorbs year, sector and sector-by-year effects; Panel B adds the 3-digit SIC absorber, which removes most of the variation in peer R&D because peer R&D is itself defined at the industry-year level. Standard errors are clustered by firm. The sample period covers 1997 to 2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Firm profit growth $\log \Pi_{f,t+k} - \log \Pi_{f,t}$	Forward Horizon (k years)				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A: as reported in Table 11					
$\overline{\text{IDE}} (\text{PC}_1)$	−0.122*** (0.011)	−0.190*** (0.019)	−0.232*** (0.025)	−0.283*** (0.031)	−0.332*** (0.037)
$\text{IDE}^{\text{tilt}} (\text{PC}_2)$	−0.002 (0.002)	−0.001 (0.003)	0.000 (0.005)	−0.005 (0.006)	−0.004 (0.007)
$\overline{\text{IDE}} (\text{PC}_1) \times \text{Product Market Similarity (log)}$	−0.013*** (0.004)	−0.018*** (0.006)	−0.020** (0.009)	−0.030*** (0.009)	−0.039*** (0.011)
$\text{IDE}^{\text{tilt}} (\text{PC}_2) \times \text{Product Market Similarity (log)}$	0.000 (0.001)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	−0.001 (0.003)
Observations	49,519	45,496	42,079	39,154	36,575
R ²	0.134	0.152	0.148	0.147	0.153
Panel B: every control additionally interacted with the moderator					
$\overline{\text{IDE}} (\text{PC}_1)$	−0.122*** (0.011)	−0.188*** (0.019)	−0.229*** (0.025)	−0.279*** (0.031)	−0.328*** (0.037)
$\text{IDE}^{\text{tilt}} (\text{PC}_2)$	−0.002 (0.002)	−0.001 (0.003)	0.000 (0.005)	−0.005 (0.006)	−0.004 (0.007)
$\overline{\text{IDE}} (\text{PC}_1) \times \text{Product Market Similarity (log)}$	−0.014*** (0.005)	−0.019*** (0.007)	−0.020** (0.009)	−0.031*** (0.009)	−0.040*** (0.011)
$\text{IDE}^{\text{tilt}} (\text{PC}_2) \times \text{Product Market Similarity (log)}$	0.000 (0.001)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	−0.001 (0.003)
Observations	49,519	45,496	42,079	39,154	36,575
R ²	0.135	0.152	0.149	0.148	0.154

TABLE E.4: Robustness of the product-market-similarity moderator to interacting the controls. Panel A reproduces Equation (22) as reported in Table 11. Panel B adds an interaction between the moderator and each element of the control vector. Both panels include sector-by-year and 3-digit SIC industry fixed effects, with standard errors clustered at the firm level, and are estimated on the same firm-years. The data are from 1997 to 2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

E.4 The annual distribution of composite IDE

Figure E.1 shows the full annual distribution of composite IDE as a violin plot. The annual (within-year) distributions are mildly left-skewed in almost every year, with a tail of firms at low exposure; the pronounced right skew of the pooled distribution documented in Section 3.1 instead reflects variation in the annual means across years, most visibly the elevated early-2000s cross-sections. The annual medians (red diamonds) trace the same wave-like path as the revenue-weighted mean—elevated around the late-1990s technology boom, lower through the post-crisis years, and recovering in the late 2010s.

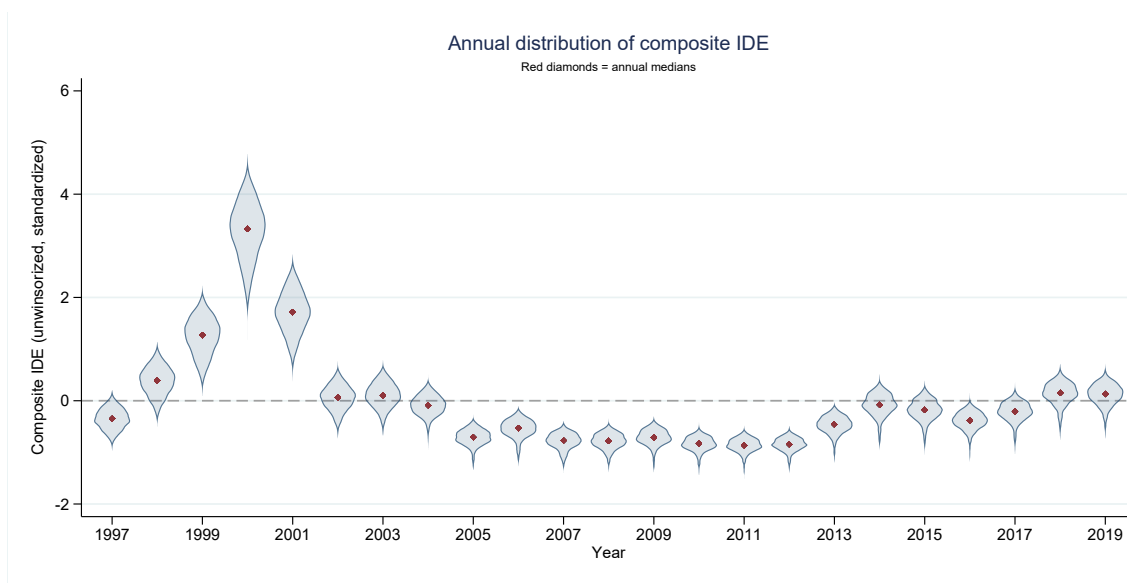


Figure E.1: Annual distribution of composite IDE, 1997–2019. Each violin shows the cross-sectional distribution of standardized composite IDE ($\overline{IDE}$) in a given year; red diamonds mark annual medians.