Measuring Creative Destruction[†]

Ali Kakhbod UC Berkeley

Leonid Kogan MIT, NBER

Peiyao Li UC Berkeley

Dimitris Papanikolaou Northwestern, NBER

May 12, 2026

Abstract

We construct a firm-level measure of creative destruction based on the similarity between the focal firm's technology description from 10k document and patented innovation of other firms. Motivated by a simple model, we identify two channels of displacement, one operating in the product market and one through technologies the firm uses in production, and build a composite measure that summarizes both. Variation in our composite displacement measure is associated with significant declines in firm profits, as well as declines in revenue, employment, and capital. We find no link between creative destruction at the industry level and subsequent growth, but we do see significant reallocation: industries with higher levels of 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 November 2025. 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 second tradition—Akcigit and Ates (2019); Davis and Haltiwanger (1992); Haltiwanger (2012); Kogan, Papanikolaou and Stoffman (2020), with Acemoglu et al. (2018) integrating the two views in a single framework—reads the same process as primarily a reallocation mechanism: the most visible margins of creative destruction are employment churn and the widening of the cross-firm distribution of growth, while the implications for the average level of industry productivity are less clear-cut. Adjudicating between these views 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 aging of a firm's own patents (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). None of these measures directly summarizes 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 are to the focal firm's existing technologies, weighted by the economic value of those innovations. A simple 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 percent at one year and 29 percent 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 adjudicates between the two views with which the paper opens. Composite displacement is small and statistically indistinguishable from zero for average industry productivity in both the Compustat sum-of-sums series and 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 0.564. 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 the Product Market Similarity measure of 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 attenuates toward zero. 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 adjudicating between 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; Akcigit and Ates, 2019; Davis and Haltiwanger, 1992; Kogan, Papanikolaou and Stoffman, 2020). 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. 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 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 (full derivation in Appendix D) in which external innovations reduce the focal firm's profits through two related channels. A prominent feature of the setup is that the same location of rival innovation in technology space maps simultaneously into product-market and factor-market incentives, yet the transmission runs through distinct equilibrium objects. In the *product channel*, rivals' quality improvements shift demand away from the focal firm. In the *process channel*, advances in processes the firm uses tighten the corresponding factor markets and raise its costs. The two channels share the same underlying force—innovation close to what the firm does—but operate through output prices versus factor prices. Both contribute to a single composite displacement effect on profits, and both rationalize factor-specific responses on the firm's factor demand once generalized beyond the profit accounting alone.

By a *process* we mean a specific technology that the firm uses in production—a method, technique, or capability deployed to make, deliver, or operate. Two firms producing similar products can rely on very different processes. Integrated steel mills smelt iron ore in basic-oxygen furnaces; mini-mills melt scrap in electric-arc furnaces; both ship rolled steel, but their process stacks—the bundle of furnaces, raw-material pipelines, and skilled labor each draws on—are nearly disjoint. Brick-and-mortar retailers organize selling around stores, in-store inventory, and store labor; e-commerce retailers organize the same activity around fulfillment centers, automation, and last-mile logistics. The model below treats each firm as drawing on a basket of such processes. External innovations close to those processes tighten the corresponding factor markets and raise marginal costs; innovations close to the firm's products erode its quality-adjusted prices.

Demand.—Within each product market m, a continuum of firms competes under monopolistic CES preferences with elasticity of substitution $\theta_m > 1$. The focal firm's revenue share depends on its quality-adjusted price relative to rivals (Appendix D.2).

Production.—Each firm produces using multiple processes $j \in J$, combined via CES 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 $\alpha \in (0, 1)$ (Appendix D.3).

Factor markets.—Labor is specific to each process j and clears across firms and products. Aggregate capability in process j is $K(j) \equiv \sum_f k(f, j)$. External innovations by other firms increase $K(j)$, which tightens the process-j labor market and raises the equilibrium wage $w(j)$, increasing costs for every firm that relies on that process.

Pricing.—Constant-markup pricing under CES implies that proportional changes in profits equal proportional changes in revenue (Appendix D.4).

Product channel.—When a rival g in market m improves its product quality ($d \log Q_{gm} > 0$), its quality-adjusted price $\tilde{p}_{gm}$ falls. Under CES demand, expenditure shifts toward g and away from the focal firm f. Under the benchmark of fixed within-market expenditure ($d \log E_m = 0$), log-differentiating the focal firm's revenue yields:

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

Product displacement is a share-weighted sum of rivals' quality improvements, scaled by $(\theta_m - 1)$: it is stronger when demand is more elastic and when quality growth is concentrated among large rivals. Aggregating across the firm's product markets with profit weights $\pi_{fm} \equiv \Pi_{fm}/\Pi_f$ yields the firm-level product channel:

$$d \log \Pi_f^{\text{prod}} = -\sum_{m \in M_f} \pi_{fm} \, (\theta_m - 1) \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} \, d \log Q_{gm}. \tag{2}$$

The firm's total product-channel profit decline is a profit-weighted average of within-market displacement, with markets that contribute more to total profit receiving greater weight. The full derivation is in Appendix D.5.

Process channel.—When external innovations raise aggregate capability $K(j)$ in process j, the process-j labor market tightens and wages rise, raising costs for every firm that relies on that process. Let $s_{f,j}$ denote the focal firm's *technology exposure share*—the fraction of its effective capability drawn from process j—let $\bar{s}_{m,j} \equiv \sum_{g \in \mathcal{F}_m} \sigma_{gm} \, s_{g,j}$ denote the expenditure-weighted average exposure of firms in market m, and let ω_j denote process j's economy-wide capability share. The product-level elasticity of profits to a process-j capability increase is:

$$\begin{aligned} \frac{\partial \log \Pi_{fm}^{\text{proc}}}{\partial \log K(j)} = & \underbrace{-\frac{\psi - 1}{\psi}(1-\alpha)\left[1 + \frac{(\theta_m - \psi)(\theta_m - 1)}{\theta_m}\right] s_{f,j}}_{\text{direct cost effect}} \\ & + \underbrace{\frac{\psi - 1}{\psi}(1-\alpha)\frac{(\theta_m - \psi)(\theta_m - 1)}{\theta_m}\,\bar{s}_{m,j}}_{\text{within-market reallocation}} \; \underbrace{-\frac{\psi - 1}{\psi}\alpha\,\omega_j}_{\text{economy-wide GE}} \; . \end{aligned} \tag{3}$$

The first term is a *direct cost effect*: the firm's marginal cost rises in proportion to its own exposure $s_{f,j}$. The second is a *within-market reallocation* effect: because competitors in market m also face higher costs, the net displacement depends on exposure relative to the market average $\bar{s}_{m,j}$. The third is an *economy-wide general-equilibrium* correction operating through cross-process substitution, proportional to ω_j. The formal definitions of $s_{f,j}$, $\bar{s}_{m,j}$, and ω_j and the full derivation are in Appendix D.6. Aggregating across the firm's product markets with profit weights π_{fm} and across processes weighted by innovation arrivals $\mu(j)\,dN_{j,t}$ yields the firm-level process channel:

$$d\log\Pi_f^{\text{proc}} = \sum_{j\in J} \underbrace{\left[\sum_{m\in M_f} \pi_{fm}\,\frac{\partial \log \Pi_{fm}^{\text{proc}}}{\partial \log K(j)}\right]}_{\text{firm-level elasticity to process } j} \mu(j)\,dN_{j,t}. \tag{4}$$

The firm's total process-channel profit decline sums across processes: for each j, the firm-level profit elasticity (itself a profit-weighted average of the product-level elasticities) is multiplied by the innovation arrival in that process.

Implications for measurement (common structure).—The firm-level product channel (2) and process channel (4) share three ingredients: (i) the *relevance* of other firms' innovations to the focal firm's technologies—through within-market shares in the product channel and exposure shares in the process channel; (ii) the *economic magnitude* of those innovations; and (iii) a *technology-specific match*, with the product channel running through technologies the firm offers in its markets and the process channel running through technologies the firm uses internally. Total profit displacement equals the sum of the two channels:

$$d\log\Pi_f \;=\; d\log\Pi_f^{\text{prod}} + d\log\Pi_f^{\text{proc}}. \tag{5}$$

We construct an empirical analog for each channel separately and summarize their joint variation. The two channels share underlying ingredients but differ on the technology-specific match, so the channel-specific measures comove tightly in levels and more loosely within sector-by-year cells; Section 2.3 documents the correlation structure and summarizes $(\text{IDE}^{\text{process}}, \text{IDE}^{\text{product}})$ via principal component analysis as a composite level of displacement and an orthogonal asymmetry component, each with its own economic content.

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 expenses, employment, total assets, capital stock, market capitalization, and 3-digit SIC industry codes). These datasets are merged using the PERMNO firm identifier and 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, 67,050 firm-years carry valid IDE scores; the largest balanced sample for the one-year profit-growth specification in Table 4 is 50,028 firm-years at horizon $t+1$. 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 average one-year profit growth in the sample is 3.3 percent and the average five-year profit growth is 15.3 percent.

2.3 Empirical construction

We now describe how we map the theoretical objects summarized in Equations (2) and (4) 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 j'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.6; Appendix A.7 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 $\mathbf{Innov}_{j,t}$ using a text embedding model (text-embedding-3-large, producing 3072-dimensional embeddings).

Technology stack. We characterize firm i'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.

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. Both stages run at temperature zero. 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—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.4; Appendix A.5 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: $\mathbf{Tech}_{i,t}^{\text{process}}$ and $\mathbf{Tech}_{i,t}^{\text{product}}$.

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).

Insert Table 1 and 2 here.

Tables 1 and 2 provide intuitive evidence that this decomposition captures distinct dimensions of the firm. For Microsoft, the process summary is dominated by terms related to internal operations and R&D (testing, facilities, manufacturing, scientific work), while the product summary emphasizes consumer-facing products (Word, search, advertising, live services). For Coca-Cola, the process summary captures upstream operations (raw materials, commodities, delivery logistics, management), while the product summary captures brands and products (energy drinks, Monster, Dasani). 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 i in year t, we compute:

$$
\begin{aligned}
IDE_{i,t}^{\text{process}} &= \sum_{j\neq i}^{n_{t,\text{Innov}}} \cos(\mathbf{Innov}_{j,t}, \mathbf{Tech}_{i,t}^{\text{process}}) \cdot A_{j,t}^{f}, \\
IDE_{i,t}^{\text{product}} &= \sum_{j\neq i}^{n_{t,\text{Innov}}} \cos(\mathbf{Innov}_{j,t}, \mathbf{Tech}_{i,t}^{\text{product}}) \cdot A_{j,t}^{f}, \qquad (6)
\end{aligned}
$$

where $n_{t,\text{Innov}}$ is the number of innovative firms in year t and $A^f_{j,t}$ is the aggregate innovation value (relative to size) of firm j in year t from Kogan et al. (2017). Here $A^f_{j,t}$ 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. Product IDE approximates Equation (2): the cosine similarity between firm j's innovation embedding and firm i's product-technology embedding replaces the product of the market share σ_{gm} and quality improvement $d\log Q_{gm}$, while the innovation value $A^f_{j,t}$ captures the economic magnitude of the innovation. Process IDE approximates Equation (4): the cosine similarity to the process-technology embedding replaces the innovation arrival $\mu(j)\,dN_{j,t}$, again weighted by $A^f_{j,t}$.

$\text{IDE}^{\text{product}}_{i,t}$ 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 market. $\text{IDE}^{\text{process}}_{i,t}$ 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 variable costs (Compustat: cogs); log of the firm's capital stock (Compustat: ppegt); 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, $A^f_{f,t}$; firm age (years since the firm first appears in Compustat); 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.4), 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

$$s_{fg,t} \;\equiv\; \frac{\mathbf{p}_{f,t}^{\top}\mathbf{p}_{g,t}}{\|\mathbf{p}_{f,t}\|\,\|\mathbf{p}_{g,t}\|}. \tag{7}$$

The *raw* product-market similarity sum is then

$$\mathrm{PMS}_{f,t} \equiv \sum_{g\in\mathcal{F}_t\setminus\{f\}} s_{fg,t}. \tag{8}$$

Higher values of $\mathrm{PMS}_{f,t}$ 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 $\log \mathrm{PMS}_{f,t}$, 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, so the embedding model, text span, and vector normalization match Tables 3 and 9.

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

$$\mathrm{PWSG}_{f,t} \equiv \frac{\sum_{g\in\mathcal{F}_t\setminus\{f\}} s_{fg,t}\left[\log \mathrm{sale}_{g,t} - \log \mathrm{sale}_{g,t-1}\right]}{\sum_{g\in\mathcal{F}_t\setminus\{f\}} s_{fg,t}}. \tag{9}$$

High values of $\mathrm{PWSG}_{f,t}$ 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 $\mathrm{IDE}^{\mathrm{process}}$ and $\mathrm{IDE}^{\mathrm{product}}$. We regress each standardized channel-specific measure on a panel of firm characteristics, including 3-digit SIC industry fixed effects and Fama–French 30 sector-by-year fixed effects, 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 the multivariate regression are firm age, firm size (capital stock), current profitability, and the level of product market crowdedness ($\mathrm{PMS}_{f,t}$). 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.

Taken together, the observable firm characteristics in Table 3 explain about 12 percent of the 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 67,050 firm-years for which both are defined, the raw Pearson correlation between $\mathrm{IDE}^{\mathrm{process}}_{f,t}$ and $\mathrm{IDE}^{\mathrm{product}}_{f,t}$ is 0.99, but much of that comovement reflects comovement across fiscal years and across coarse industries. To capture the residual variation, we demean $\mathrm{IDE}^{\mathrm{process}}$ and $\mathrm{IDE}^{\mathrm{product}}$ 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 $(\mathrm{IDE}^{\mathrm{process}}_{f,t}, \mathrm{IDE}^{\mathrm{product}}_{f,t})$ over the full firm-year sample and defining our two summary measures as the standardized first and second principal components:

$$\overline{\mathrm{IDE}}_{f,t} \equiv \widetilde{\mathrm{PC}}_{1,f,t}, \qquad \mathrm{IDE}^{\mathrm{tilt}}_{f,t} \equiv \widetilde{\mathrm{PC}}_{2,f,t}. \tag{10}$$

We sign $\overline{\mathrm{IDE}}$ so that its loadings on $\mathrm{IDE}^{\mathrm{process}}$ and $\mathrm{IDE}^{\mathrm{product}}$ are jointly positive (here, equal at 0.71), so that higher $\overline{\mathrm{IDE}}$ unambiguously means more displacement on *both* sides; we sign $\mathrm{IDE}^{\mathrm{tilt}}$ so that its loading on $\mathrm{IDE}^{\mathrm{process}}$ is positive (and on $\mathrm{IDE}^{\mathrm{product}}$ negative), so that positive $\mathrm{IDE}^{\mathrm{tilt}}$ 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 $\mathrm{IDE}^{\mathrm{tilt}}$ 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 cumulative changes 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

Insert Figure 3 here.

Figure 3 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 2.3, which splits the joint within-FF30 variance of $(\text{IDE}^{\text{process}}, \text{IDE}^{\text{product}})$ into a symmetric direction and an orthogonal tilt. Both decompositions point in the same practical direction: the bulk of the firm-year dispersion in composite IDE is within Fama–French 30 industry-by-year cells, 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 ascending 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.

Appendix Figure 2 shows the distribution of firms across sectors. The largest sectors are Personal and Business Services, Healthcare/Pharmaceutical Products, and Business Equipment (600–1,300 firms each). The smallest are Textiles and Tobacco Products (fewer than 50 firms each).

3.2 Time-series variation

We next examine the time-series properties of composite IDE. 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. We omit 1998 from the figure because of sparse patent coverage in that year and rebase the index so that 1997 remains 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.

Insert Figure 4 here.

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* (entering minus exiting firms). The within component accounts for the bulk of the cumulative change from 1997 through 2019, while between and net entry/exit jointly make a modest offsetting contribution. Thus time variation in aggregate exposure primarily reflects changes in continuing incumbents' positions in the technology space, not turnover in the Compustat 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. The decomposition shows that most of the cumulative movement in the revenue-weighted mean reflects within-firm changes in measured exposure rather than reallocation across firms or net entry and exit, which is 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 $+0.36$ standard deviations versus a sector average of firm means near $+0.04$. 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 ($+0.29$ 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 68th percentile in consumer goods ($+0.17$ versus a sector mean of firm means of $+0.12$). Its within-sector percentile remains in the top decile across multi-year windows (roughly the 81st to the 90th 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 81st 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. Post-reorganization years show lower relative exposure as the firm shrinks and pivots.

Ford Motor Co. Ford's full-sample composite IDE is near the sector median (50th percentile), masking pronounced dynamics in the EV transition. Ford's within-sector percentile on composite IDE rises from roughly the 39th 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 (48th percentile), with a firm mean slightly below the sector average of firm means. Sub-period percentiles start near the 47th 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 38th percentile among business-equipment firms on average, with sub-period percentiles between roughly the 12th and the 37th, 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 (21st 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 3 documents that the bulk of the variation in composite IDE is within Fama–French 30 industries; the regressions below exploit this within-industry dispersion (in addition to year variation) via sector-by-year and industry fixed effects. 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:

$$\log \Pi_{f,t+k} - \log \Pi_{f,t} = \beta_1^k \, \overline{\mathrm{IDE}}_{f,t} + \beta_2^k \, \mathrm{IDE}^{\mathrm{tilt}}_{f,t} + \boldsymbol{\lambda}^{k\top} \mathbf{X}_{f,t} + \delta^k_{s\times t} + \gamma^k_g + \epsilon^k_{f,t}, \tag{11}$$

where the dependent variable is the change in log profits for horizons $k \in \{1, \ldots, 5\}$. 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 $\text{IDE}^{\text{tilt}}_{f,t}$ 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 (9)), using the cosine similarities $s_{fg,t}$ from Equation (7) 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^k_g and sector-by-year fixed effects $\delta^k_{s\times t}$. 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 double-clustered by firm and year.

Insert Table 4 here.

Table 4 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 percent lower profit growth rate over the next year and a 29.0 percent 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 (3.3 percent), 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.

4.2 Broader growth outcomes

We next estimate regressions analogous to Equation (11) 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 $\text{IDE}^{\text{process}}$ and $\text{IDE}^{\text{product}}$, 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 other firms in the same 3-digit SIC industry; intangible capital follows L. Eisfeldt, T. Kim and Papanikolaou (2022).

Insert Tables 5 and 6 here.

Composite displacement and top-line outcomes. Table 5, columns (1)–(5), and Table 6, columns (1)–(5), report the analogue of Equation (11) 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 percent in revenue (Panel A of Table 5), 31.8 percent in market share (Panel B of Table 5), 23.9 percent in employment (Panel A of Table 6), 44.1 percent in physical capital (Panel B of Table 6), and 36.9 percent in intangible capital (Panel C of Table 6). 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 (34)), while physical and especially intangible 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 $(\text{IDE}^{\text{process}}, \text{IDE}^{\text{product}})$—non-trivial but small relative to the level component (Section 2.3). 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.

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—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.4.6 and Appendix A.4.4.

4.3.1 Specialist labor market tightening

We use the SemanticAxis approach of Fedyk et al. (2024) and 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 splst risk$_{f,t}$ (the sentence pairs are in Appendix A.4.5).

We estimate regressions of the form:

$$\text{splst risk}_{f,t+k} = \beta_1^{L,k}\,\overline{\text{IDE}}_{f,t} + \beta_2^{L,k}\,\text{IDE}^{\text{tilt}}_{f,t} + \beta_{\text{lag}}^{L,k}\,\text{splst risk}_{f,t-5} + \boldsymbol{\lambda}^{k\top}\mathbf{X}_{f,t} + \delta_{s\times t}^{L,k} + \gamma_g^{L,k} + \epsilon_{f,t}^{L,k}, \quad (12)$$

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 7 here.

Table 7 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—and slightly strengthens—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 roughly 90 percent of filings (the sentence pairs and extraction pipeline are in Appendix A.4).

We estimate analogous regressions:

$$\text{ob risk}_{f,t+k} = \beta_1^{O,k}\,\overline{\text{IDE}}_{f,t} + \beta_2^{O,k}\,\text{IDE}^{\text{tilt}}_{f,t} + \beta_{\text{lag}}^{O,k}\,\text{ob risk}_{f,t-5} + \boldsymbol{\lambda}^{k\top}\mathbf{X}_{f,t} + \delta^{O,k}_{s\times t} + \gamma^{O,k}_{g} + \epsilon^{O,k}_{f,t}, \tag{13}$$

where the specification mirrors (12) 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 8 here.

Table 8 reports the analogue for product obsolescence risk. Composite IDE enters positively at every horizon (coefficients between 0.11 and 0.26 standard deviations), 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.

Taken together, the two mechanism tables indicate that firms with high composite IDE subsequently disclose both tighter specialist labor markets and elevated 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 the text-based Product Market Similarity score of 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 own-patent aging.

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 of Hoberg and Phillips (2016) introduced in Section 2.4 (Equations (7) and (8)). 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 $\log \mathrm{PMS}_{f,t}$ and with its interactions with composite displacement and the asymmetry component:

$$\begin{aligned}\Pi_{f,t+k} - \Pi_{f,t} = \beta_1^k\, \overline{\mathrm{IDE}}_{f,t} &+ \beta_2^k\, \mathrm{IDE}^{\mathrm{tilt}}_{f,t} + \beta_3^k\, \log \mathrm{PMS}_{f,t} \\ &+ \beta_4^k\, \overline{\mathrm{IDE}}_{f,t} \times \log \mathrm{PMS}_{f,t} + \beta_5^k\, \mathrm{IDE}^{\mathrm{tilt}}_{f,t} \times \log \mathrm{PMS}_{f,t} \\ &+ \boldsymbol{\lambda}^{k\top} \mathbf{X}_{f,t} + \delta^k_{s\times t} + \gamma^k_g + \epsilon^k_{f,t}. \end{aligned} \tag{14}$$

Here, the dependent variable, the control vector $\mathbf{X}_{f,t}$, the sector-by-year fixed effects $\delta^k_{s\times t}$, and the 3-digit SIC industry fixed effects γ^k_g are identical to those in Equation (11); $\overline{\mathrm{IDE}}_{f,t}$, $\mathrm{IDE}^{\mathrm{tilt}}_{f,t}$, and $\log \mathrm{PMS}_{f,t}$ all enter standardized to mean zero and unit variance over the full sample. Standard errors are double-clustered by firm and year. 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 9 here.

Table 9 reports the estimates of Equation (14). The coefficient on composite IDE remains negative at every horizon: a one-standard-deviation increase is associated with about a 12.2 percent lower one-year profit growth rate and about a 33.2 percent lower cumulative rate by year five. The coefficient on the asymmetry component is small at every horizon (between -0.005 and 0.000), consistent with profit growth loading on the level of displacement rather than on its process-versus-product tilt, once Product Market Similarity is held fixed. Product Market Similarity on its own is small in magnitude and indistinguishable from zero at every horizon (coefficients between -0.018 and 0.008 across the five horizons). The interaction of composite IDE with Product Market Similarity is negative and grows in magnitude with the horizon (from about -0.013 at year one to about -0.039 at year five), so the cumulative profit response to a given increase in composite displacement is more negative when the firm operates in a more crowded product space. The interaction of the asymmetry component with Product Market Similarity is small at every horizon.

5.2 IDE and patent-citation-based technological obsolescence

Ma (2025) measures technological obsolescence using the decline in external citations received by a firm's patent base. This measure differs from IDE in three ways. First, it is restricted to patenting firms (roughly 30 percent of our sample). Second, it captures aging of the firm's *own* patent portfolio rather than exposure to external innovations affecting the firm's full technology base. Third, it does not distinguish process from product technologies. We use the five-year-lookback 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 10 here.

Table 10 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 attenuates toward zero once composite IDE is included. 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 the sample to industries with at least five IDE-valid firms over the sample. 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 covers the full universe of producers—public and private—and removes 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 it does not contain sufficient information for us to compute measures of reallocation—dispersion in firm growth rates.

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

$$\log Y_{i,t+k} - \log Y_{i,t} = \beta_{1,k}\, \overline{\mathrm{IDE}}_{i,t} + \beta_{2,k}\, \mathrm{IDE}^{\mathrm{tilt}}_{i,t} + \gamma_k\, Y^{m}_{i,t} + \alpha_t + \varepsilon_{i,t,k}, \tag{15}$$

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

Insert Table 11 here.

Insert Table 12 here.

Table 11 focuses on the Compustat sample while Table 12 focuses on the BEA sample. We see that, across both samples, both productivity definitions, and all horizons, composite displacement has no statistically significant relation with subsequent growth in industry productivity. The asymmetry component is also weak in absolute magnitude—marginally negative at the three- and five-year horizons in the BEA 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 $k \in \{1, \ldots, 5\}$:

$$\text{spread}_{g_{t+k};i,t} = \beta_{1,k}\,\overline{\text{IDE}}_{i,t} + \beta_{2,k}\,\text{IDE}^{\text{tilt}}_{i,t} + \gamma_k\,\text{spread}_{g_{t+k},\,i,t} + \alpha_t + \varepsilon_{i,t,k}, \tag{16}$$

where $\text{spread}_{g_{t+k};i,t}$ is the within-industry dispersion between the ninetieth and tenth percentiles of k-year firm growth rates in measured productivity—revenue or gross profits per worker—in industry i at date t. We control for the contemporaneous ninetieth-to-tenth spread of the level of log gross profit per worker $\text{spread}_{i,t}$ in industry i in year t. $\overline{\text{IDE}}_{i,t}$ and $\text{IDE}^{\text{tilt}}_{i,t}$ are the innovation displacement exposure measures transformed into industry-year percentile ranks as discussed above. As before, we include fiscal-year fixed effects α_t and cluster the errors at the industry level.

Insert Table 13 here.

Panels A and B of Table 13 reports 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, increasing the 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, rising to 0.79 at the four-year horizon. 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, creative destruction does not raise the level of industry productivity in our sample but rather leads to increased dispersion in firm outcomes.

7 Robustness

We assess the robustness of our main profit-growth results in Table 4 along four 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 lagged IDE values to current firm-years to trace how fast predictive content decays when the technology stack and innovation side are deliberately desynchronized. 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 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 lagged IDE values to current firm-years. Under the null that timing is irrelevant, both placebo measures should absorb the predictive content of correctly timed IDE; the data reject the null 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 analogues:

$$\begin{aligned}\log \Pi_{f,t+k} - \log \Pi_{f,t} = \beta_1^k \, \overline{\text{IDE}}_{f,t} + \beta_2^k \, \text{IDE}^{\text{tilt}}_{f,t} + \beta_3^k \, \overline{\text{IDE}}_{f,t}^{\text{shuffle}} + \beta_4^k \, \text{IDE}^{\text{tilt}^{\text{shuffle}}}_{f,t} \\ + \boldsymbol{\lambda}^{k\top} \mathbf{X}_{f,t} + \delta^k_{s\times t} + \gamma^k_g + \epsilon^k_{f,t}, \end{aligned} \quad (17)$$

where $\overline{\text{IDE}}_{f,t}^{\text{shuffle}}$ and $\text{IDE}^{\text{tilt}^{\text{shuffle}}}_{f,t}$ denote the composite displacement and asymmetry components formed by applying the same principal-component convention as in Section 2.3 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 analogues should absorb the variation that drives our results.

Insert Table 14 here.

Table 14 reports that composite displacement remains negative at every horizon once the shuffled-stack controls are included: a one-standard-deviation increase is associated with about an 11.8 percent decline in one-year profit growth and about a 27.6 percent decline by year five. The asymmetry component is small and statistically insignificant at every horizon. The magnitudes line up closely with Table 4, which indicates that 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 LLM-derived summaries could inadvertently encode information that is not contemporaneous with the innovation side of IDE, so that similarity to innovations could line up with subsequent outcomes for reasons unrelated to true displacement pressure at date t. To probe this, we attach to firm-year (f, t) the IDE originally measured at $(f, t-h)$ and re-estimate the baseline profit-growth regression separately for each $h \in \{1, 3, 5\}$:

$$\log \Pi_{f,t+k} - \log \Pi_{f,t} = \beta_1^k \, \overline{\text{IDE}}_{f,t-h} + \beta_2^k \, \text{IDE}^{\text{tilt}}_{f,t-h} + \boldsymbol{\lambda}^{k\top} \mathbf{X}_{f,t} + \delta^k_{s\times t} + \gamma^k_g + \epsilon^k_{f,t}, \tag{18}$$

where the lagged regressors $\overline{\text{IDE}}_{f,t-h}$ and $\text{IDE}^{\text{tilt}}_{f,t-h}$ replace their contemporaneous counterparts in Equation (11), and the control vector $\mathbf{X}_{f,t}$ and the fixed effects $\delta^k_{s\times t}$ and γ^k_g are unchanged. Because IDE is highly autocorrelated, lagged values can continue to predict profit growth, but β_1^k should shrink as h grows if the signal is anchored in contemporaneous alignment between technology stacks and innovations; conversely, if the summaries leaked future information, predictive power should rise rather than decay with h.

Insert Table 15 here.

Table 15 shows that composite displacement remains negative when attached backward by one year ($h = 1$), but the magnitudes shrink sharply once h reaches three or five years. The asymmetry component is small throughout. The decay pattern is what we expect if predictive content weakens when the technology-innovation match is deliberately desynchronized—the opposite of what we would see if the measurement systematically exploited future information in the summaries.

7.2 IDE without innovation-value weights

Our baseline IDE weights cosine similarity by other firms' innovation values (scaled by the focal firm's 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:

$$\begin{aligned} IDE^{\text{process,raw}}_{i,t} &= \sum_{j\neq i}^{n_{t,\text{Innov}}} \cos(\mathbf{Innov}_{j,t}, \mathbf{Tech}^{\text{process}}_{i,t}), \\ IDE^{\text{product,raw}}_{i,t} &= \sum_{j\neq i}^{n_{t,\text{Innov}}} \cos(\mathbf{Innov}_{j,t}, \mathbf{Tech}^{\text{product}}_{i,t}). \end{aligned} \tag{19}$$

We then form composite displacement and the asymmetry component from the standardized pair $(IDE_{i,t}^{\text{process,raw}}, IDE_{i,t}^{\text{product,raw}})$ exactly as in Section 2.5 and re-estimate Equation (11).

Insert Table 16 here.

Table 16 reports the estimates. Dropping the weights attenuates the composite coefficient—consistent with down-weighting small innovations sharpening the signal—but composite displacement remains negative at every horizon and the asymmetry component is small throughout. At the five-year horizon, the point estimate is roughly 40 percent of the magnitude in Table 4, 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 (11) to include firm age, Product Market Similarity, 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 $\overline{\text{IDE}}$ 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 18 here.

Table 18 reports the residualized specification. Residual composite displacement remains negative at every horizon and broadly matches the baseline magnitudes in Table 4; 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 (6) 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 17.

Insert Table 17 here.

Panel A includes only the high-similarity components. Composite displacement in the high tail is negative at every horizon, and the high-similarity asymmetry component is also negative at short horizons. Panel B adds the moderate-similarity components while controlling for the high-similarity components; both buckets enter negatively for composite displacement at most horizons, which indicates that displacement risk is not confined to the extreme tail alone. Panel C includes the baseline composite displacement and asymmetry component while controlling for the moderate-similarity components. Baseline composite displacement remains strongly negative throughout, while the moderate-similarity composite enters positively once the baseline is included—a pattern consistent with the moderate bucket capturing overlapping variation that the baseline composite already summarizes, leaving the baseline object to carry the incremental predictive content for profit growth.

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?"

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*.

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."

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.

Fedyk, Anastassia, Ali Kakhbod, Peiyao Li, and Ulrike Malmendier. 2024. "AI and Perception Biases in Investments: An Experimental Study." *Available at SSRN 4787249*.

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.

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." *The 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.

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.

Figures and Tables

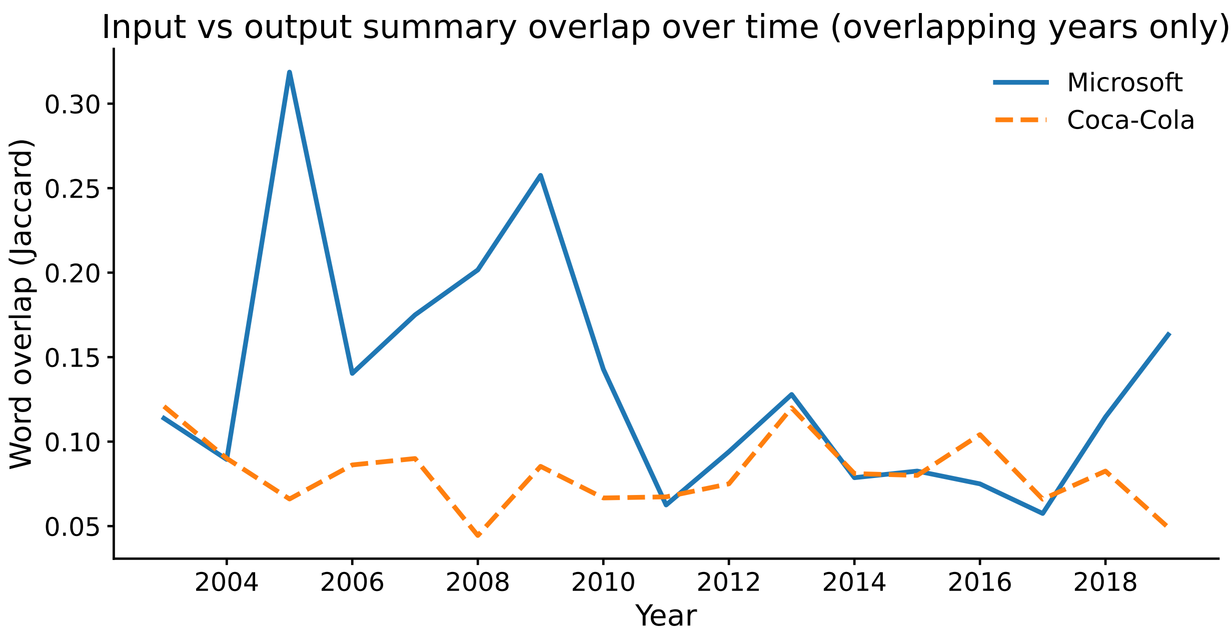

Figure 1: Token overlap between process and product technology summaries over time. For each firm-year, we plot the fraction of lemmatized tokens that appear in both summaries.

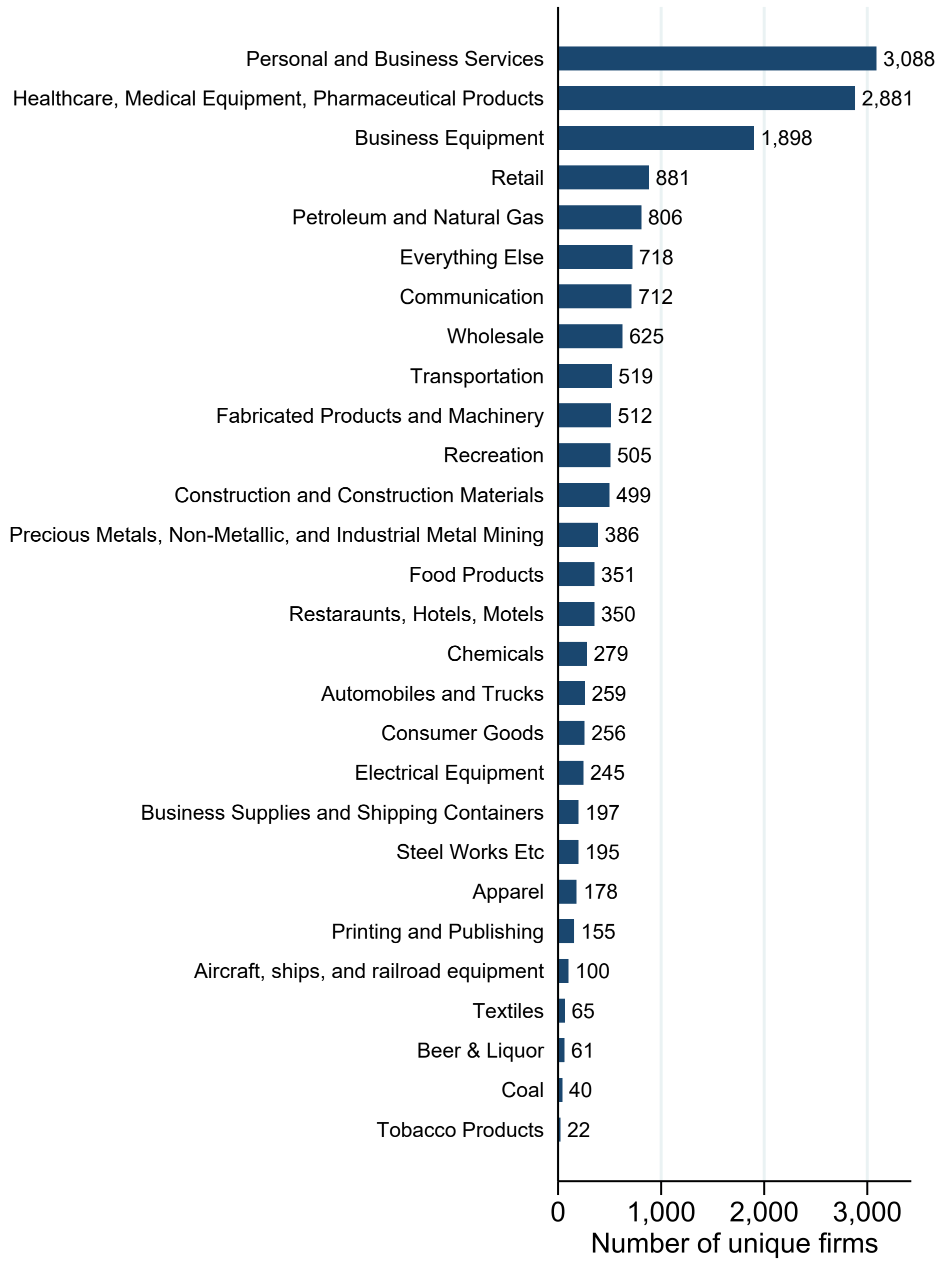

Figure 2: Number of unique firms in each sector in our sample (1997–2019).

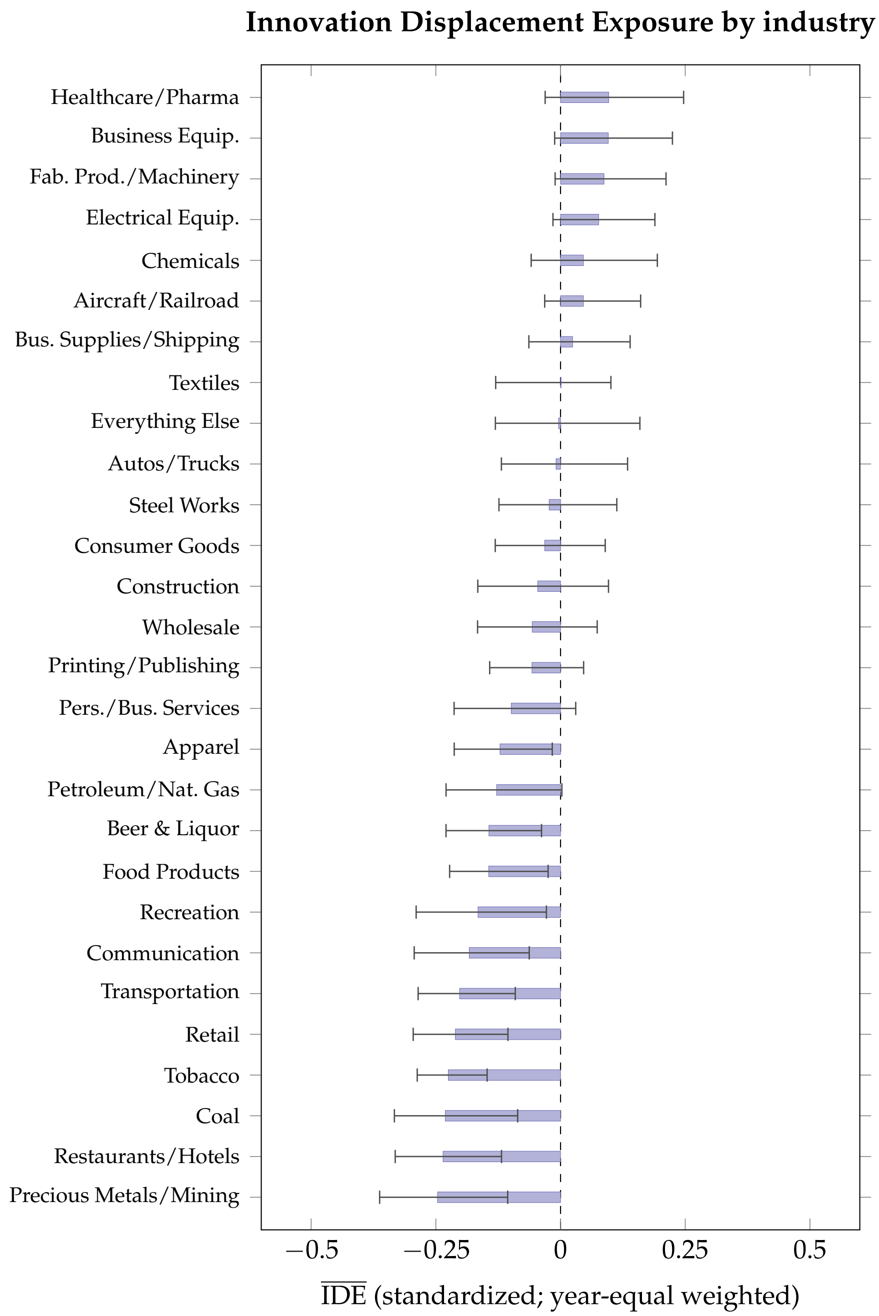

Figure 3: Innovation Displacement Exposure by industry (extended sample, 1997–2019). $\overline{\text{IDE}}$ is the standardized first principal component (PC_1) of process and product IDE. To neutralize sample-composition tilt from industries with heavy entry/exit, statistics are computed with year-equal 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 ascending order by mean (highest at top).

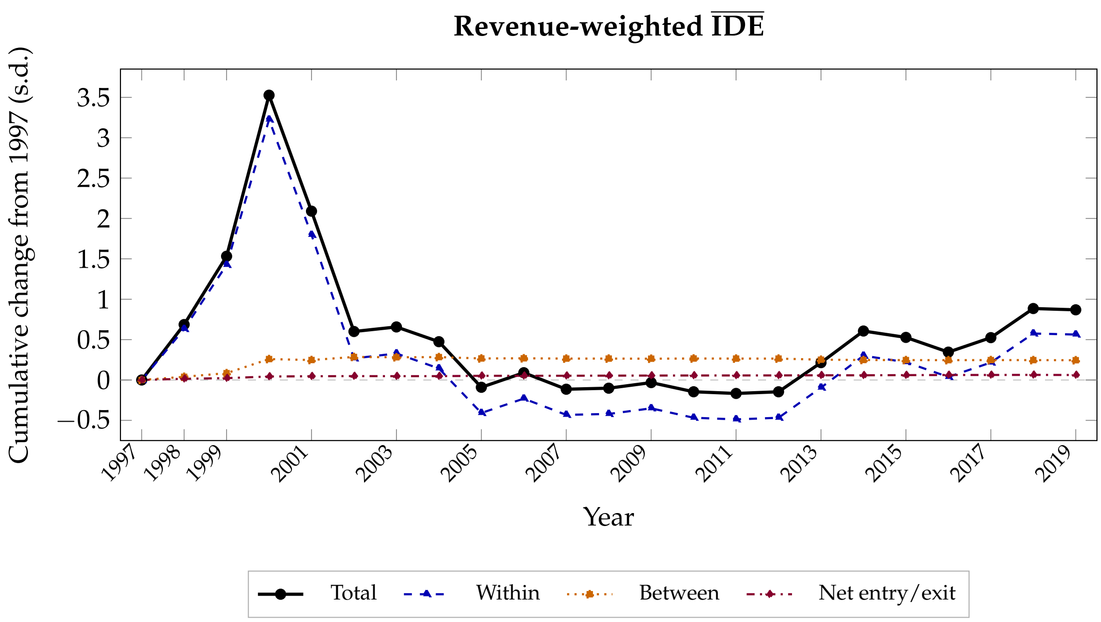

Figure 4: Decomposition of cumulative change in revenue-weighted mean $\overline{\text{IDE}}$ (1997–2019). $\overline{\text{IDE}}$ is the standardized first principal component (PC_1) of process and product IDE. The figure decomposes the cumulative change from 1997 (= 0) in the revenue-weighted cross-sectional 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)
product (22)	beverage (43)	firm (21)
process (21)	sell (40)	use (16)
procurement (15)	consumer (32)	tea (12)
assurance (14)	finish (32)	distributor (11)
improve (13)	concentrate (31)	joint (11)
internal (13)	include (31)	ready-to-drink (11)
internally (13)	syrup (31)	venture (11)
supply (13)	retailer (28)	brand (10)
information (12)	water (27)	coffee (9)
management (12)	market (26)	sparkle (9)

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^{***}	-0.180^{***}
			(0.014)	(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^{***}	0.005^{*}
			(0.002)	(0.003)
Log employment size	-0.097	-0.102	-0.005^{*}	-0.004
			(0.003)	(0.003)
Variable cost	-0.002	0.002	0.007	0.011^{**}
			(0.005)	(0.005)
Log current profit	-0.124	-0.130	0.001	0.005
			(0.003)	(0.003)
Profitability	-0.026	-0.029	0.028^{***}	0.021^{**}
			(0.008)	(0.010)
Product Market Similarity	-0.602	-0.594	0.134^{***}	0.185^{***}
			(0.009)	(0.010)
Age	-0.089	-0.090	0.008^{***}	0.008^{***}
			(0.000)	(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 double-clustered by firm and year. The sample is 1997–2019; observations with missing values in any variable are dropped. One asterisk denotes statistical significance at the 10 percent level, two at 5 percent, and three at 1 percent.

Firm profit growth	Horizon (years)				
$\log \Pi_{t+h} - \log \Pi_t$	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}}$ (PC_1)	-0.106***	-0.173***	-0.211***	-0.250***	-0.290***
	(0.011)	(0.018)	(0.024)	(0.029)	(0.035)
IDE^{tilt} (PC_2)	-0.003	-0.002	-0.000	-0.006	-0.005
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)
Firm innovation	0.027***	0.043***	0.054***	0.063***	0.073***
	(0.002)	(0.004)	(0.006)	(0.008)	(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 × Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓

TABLE 4: IDE and firm profitability. This table shows the association between our IDE measure and profit growth over the next five years (one to five years ahead). $\overline{\text{IDE}}$ (PC_1) is the composite displacement and IDE^{tilt} (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 ppegt, 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 the pairwise product similarities from 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 double-clustered at the firm and year level. The sample period covers 1997 to 2019.

	A. Revenue				
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}}$ (PC_1)	-0.112***	-0.174***	-0.221***	-0.268***	-0.308***
	(0.010)	(0.018)	(0.025)	(0.031)	(0.036)
IDE^{tilt} (PC_2)	-0.001	-0.001	-0.003	-0.006	-0.008
	(0.002)	(0.003)	(0.004)	(0.005)	(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***	-0.165***	-0.211***	-0.264***	-0.318***
	(0.011)	(0.019)	(0.026)	(0.032)	(0.038)
IDE^{tilt} (PC_2)	-0.000	-0.000	0.000	-0.002	-0.003
	(0.002)	(0.003)	(0.005)	(0.006)	(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 × Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓

TABLE 5: IDE and firm revenues and market share. This table shows the association between IDE measures and firm growth outcomes other than profit growth. $\overline{\text{IDE}}$ (PC_1) is the composite displacement and IDE^{tilt} (PC_2) 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 ppegt, 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 double-clustered at the firm and year level. The sample period covers 1997 to 2019.

	A. Firm Employment				
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}}$ (PC_1)	-0.075***	-0.127***	-0.166***	-0.200***	-0.239***
	(0.008)	(0.014)	(0.020)	(0.025)	(0.030)
IDE^{tilt} (PC_2)	-0.001	-0.001	-0.002	-0.004	-0.005
	(0.001)	(0.003)	(0.004)	(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)
$\overline{\text{IDE}}$ (PC_1)	-0.146***	-0.234***	-0.305***	-0.371***	-0.441***
	(0.008)	(0.015)	(0.021)	(0.028)	(0.033)
IDE^{tilt} (PC_2)	-0.003*	-0.004	-0.006	-0.009*	-0.010*
	(0.001)	(0.003)	(0.004)	(0.005)	(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)
$\overline{\text{IDE}}$ (PC_1)	-0.110***	-0.192***	-0.264***	-0.323***	-0.369***
	(0.005)	(0.010)	(0.016)	(0.021)	(0.026)
IDE^{tilt} (PC_2)	-0.005***	-0.008***	-0.011***	-0.013***	-0.013**
	(0.001)	(0.002)	(0.003)	(0.004)	(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 × Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓

TABLE 6: IDE and firm factor demand. This table shows the association between IDE measures and firm growth outcomes other than profit growth. $\overline{\text{IDE}}$ (PC_1) is the composite displacement and IDE^{tilt} (PC_2) 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 ppegt, 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 double-clustered at the firm and year level. The sample period covers 1997 to 2019.

Dependent variable:	Standard controls			+Lagged (5yrs) Risk Score		
Specialist market tightening risk	splst risk $t+1$	splst risk $t+2$	splst risk $t+3$	splst risk $t+1$	splst risk $t+2$	splst risk $t+3$
	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{\text{IDE}}$ (PC_1)	0.230***	0.235***	0.207***	0.275***	0.297***	0.275***
	(0.039)	(0.042)	(0.043)	(0.071)	(0.077)	(0.082)
IDE^{tilt} (PC_2)	-0.011	-0.012	-0.010	-0.001	-0.004	-0.007
	(0.008)	(0.008)	(0.009)	(0.010)	(0.011)	(0.011)
Controls						
Year × 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 7: This table shows the association between IDE and future specialist labor market tightening risk over horizons of one to three years ahead. $\overline{\text{IDE}}$ (PC_1) is the composite displacement and IDE^{tilt} (PC_2) 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. 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 ppegt, in logs), current employment (Compustat emp, 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 double-clustered at the firm and year level. The sample period covers 1997 to 2019.

Dependent variable:	Standard controls			+Lagged (5yrs) Risk Score		
Product obsolescence risk	ob risk $t+1$	ob risk $t+2$	ob risk $t+3$	ob risk $t+1$	ob risk $t+2$	ob risk $t+3$
	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{\text{IDE}}$ (PC_1)	0.162***	0.122***	0.112***	0.260***	0.232***	0.255***
	(0.032)	(0.035)	(0.036)	(0.065)	(0.071)	(0.073)
IDE^{tilt} (PC_2)	-0.001	-0.007	-0.002	0.001	-0.005	0.006
	(0.006)	(0.007)	(0.007)	(0.009)	(0.010)	(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 8: This table shows the association between IDE and future product obsolescence risk over horizons of one to three years ahead. $\overline{\text{IDE}}$ (PC_1) is the composite displacement and IDE^{tilt} (PC_2) 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. 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 ppegt, in logs), 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 double-clustered at the firm and year level. The sample period covers 1997 to 2019.

	Dependent variable:				
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{IDE}}$ (PC_1)	-0.122^{***}	-0.190^{***}	-0.232^{***}	-0.283^{***}	-0.332^{***}
	(0.011)	(0.019)	(0.025)	(0.031)	(0.037)
IDE^{tilt} (PC_2)	-0.002	-0.001	0.000	-0.005	-0.004
	(0.002)	(0.003)	(0.005)	(0.006)	(0.007)
Product Market Similarity (log)	0.004	-0.009	-0.018	-0.013	0.008
	(0.010)	(0.016)	(0.021)	(0.025)	(0.030)
$\overline{\text{IDE}}$ (PC_1) × Product Market Similarity (log)	-0.013^{***}	-0.018^{***}	-0.020^{**}	-0.030^{***}	-0.039^{***}
	(0.004)	(0.006)	(0.009)	(0.009)	(0.011)
IDE^{tilt} (PC_2) × Product Market Similarity (log)	0.000	0.001	0.001	0.001	-0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Firm innovation value	0.028^{***}	0.044^{***}	0.057^{***}	0.066^{***}	0.075^{***}
	(0.002)	(0.005)	(0.006)	(0.008)	(0.009)
Year × 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 9: Profit growth, IDE, and Product Market Similarity. $\overline{\text{IDE}}$ (PC_1) is the composite displacement and IDE^{tilt} (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 $\overline{\text{IDE}}$× Product Market Similarity and IDE^{tilt}× Product Market Similarity test whether the profit-growth effects 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 ppegt, 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 double-clustered at the firm and year level.

Firm profit growth	Forward Horizon (k years)				
$\log \Pi_{t+h} - \log \Pi_t$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A					
$\overline{\text{IDE}}$ (PC$_1$)	-0.079***	-0.134***	-0.175***	-0.211***	-0.275***
	(0.022)	(0.035)	(0.048)	(0.060)	(0.072)
IDE$^{\text{tilt}}$ (PC$_2$)	-0.005	-0.004	0.004	0.002	0.007
	(0.003)	(0.005)	(0.008)	(0.009)	(0.011)
Firm innovation	0.028***	0.045***	0.059***	0.069***	0.078***
	(0.003)	(0.005)	(0.008)	(0.009)	(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:					
$\overline{\text{IDE}}$ (PC$_1$)	-0.078***	-0.133***	-0.174***	-0.209***	-0.273***
	(0.022)	(0.035)	(0.047)	(0.059)	(0.072)
IDE$^{\text{tilt}}$ (PC$_2$)	-0.005	-0.004	0.004	0.002	0.007
	(0.003)	(0.005)	(0.008)	(0.009)	(0.011)
Technological obsolescence	-0.008**	-0.008	-0.007	-0.008	-0.009
	(0.004)	(0.006)	(0.008)	(0.010)	(0.012)
Firm innovation	0.028***	0.045***	0.059***	0.069***	0.078***
	(0.003)	(0.005)	(0.008)	(0.009)	(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 × Sector	✓	✓	✓	✓	✓

TABLE 10: 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. $\overline{\text{IDE}}$ (PC$_1$) is the composite displacement and IDE$^{\text{tilt}}$ (PC$_2$) is the asymmetry component (the first and second principal components, respectively, of process and product IDE). We include full controls: Product Market Similarity (the total product-similarity sum); 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 ppegt, 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 double-clustered at the firm and year level.

Industry Growth	Forward horizon (k years)				
(Compustat)	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A: Revenue per employee, $\Delta_k \log(\sum \text{sale} / \sum \text{emp})$					
$\overline{\text{IDE}}$	0.038	0.031	0.009	−0.022	−0.021
	(0.045)	(0.058)	(0.066)	(0.078)	(0.093)
$\overline{\text{IDE}}^{\text{tilt}}$	0.007	−0.007	−0.012	−0.006	−0.017
	(0.011)	(0.018)	(0.020)	(0.025)	(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, $\Delta_k \log((\sum \text{sale} - \sum \text{cogs}) / \sum \text{emp})$					
$\overline{\text{IDE}}$	0.067	0.036	−0.016	−0.048	−0.002
	(0.060)	(0.074)	(0.100)	(0.109)	(0.127)
$\overline{\text{IDE}}^{\text{tilt}}$	0.024	0.015	0.007	0.017	0.008
	(0.017)	(0.026)	(0.029)	(0.035)	(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 11: The table reports two definitions of industry-level labor productivity, each formed as a sum-of-sums ratio across IDE-valid 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, $\overline{\text{IDE}}^{\text{tilt}}$, 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 productivity log ratio enters as a control. We include calendar year fixed effects in the specification. Standard errors are clustered at the industry level. The data covers 1997 to 2019.

Industry Growth	Forward horizon (k years)				
(BEA)	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\overline{\text{IDE}}$	$+0.018$	$+0.048$	$+0.072$	$+0.061$	$+0.088$
	(0.020)	(0.035)	(0.049)	(0.064)	(0.076)
$\overline{\text{IDE}}^{\text{tilt}}$	-0.007	-0.007	-0.015	-0.021	-0.028
	(0.006)	(0.010)	(0.013)	(0.016)	(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 12: 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, $\overline{\text{IDE}}^{\text{tilt}}$, are the first and second principal components of the asset-weighted industry means of $\text{IDE}^{\text{process}}$ and $\text{IDE}^{\text{product}}$. All three 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. 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% on the full Compustat panel before aggregation. The data covers 1997 to 2019.

Industry Reallocation:	Forward horizon (k years)				
Dispersion in Firm Growth Rates	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Panel A: Growth in Revenue per employee, $\Delta_k \log(\text{sale}/\text{emp})_f$					
$\overline{\text{IDE}}$ (PC$_1$)	0.571***	0.649***	0.659**	0.787**	0.671*
	(0.172)	(0.232)	(0.277)	(0.324)	(0.354)
IDE$^{\text{tilt}}$ (PC$_2$)	0.030	0.008	−0.001	−0.004	−0.005
	(0.029)	(0.036)	(0.041)	(0.046)	(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***	0.355**	0.331*	0.377*	0.403*
	(0.128)	(0.164)	(0.193)	(0.219)	(0.244)
IDE$^{\text{tilt}}$ (PC$_2$)	0.035	0.013	−0.011	−0.001	0.032
	(0.033)	(0.045)	(0.052)	(0.057)	(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 13: This table reports industry-level dispersion in firm growth rates and performance or liquidation exits 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. Panel C reports Poisson regressions in which the dependent variable is the count of performance or liquidation exits among the set of firms present in industry i in year t, where performance or liquidation exits are defined by DLSTCD $\geq$ 400; the base set is fixed at year t and includes firms without IDE scores, and the exposure offset is $\log(n_{i,t})$, so coefficients approximate log-rate-ratios. The common component, $\overline{\text{IDE}}$ (PC$_1$), and the tilt component, IDE$^{\text{tilt}}$ (PC$_2$), 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; Panel C controls for $\log(1 + \text{exits in year } t)$ for the same exit type. Year fixed effects are absorbed. Standard errors are clustered at the industry level. We winsorize all firm-level variables at the 1/99% in each year. The data covers 1997 to 2019.

Firm profit growth	Forward Horizon (k years)				
$\log \Pi_{t+h} - \log \Pi_t$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\overline{\text{IDE}}$ (PC_1)	-0.118^{***}	-0.177^{***}	-0.209^{***}	-0.240^{***}	-0.276^{***}
	(0.011)	(0.018)	(0.024)	(0.029)	(0.034)
IDE^{tilt} (PC_2)	-0.004^{*}	-0.003	-0.001	-0.006	-0.004
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)
Firm innovation value	0.027^{***}	0.043^{***}	0.054^{***}	0.063^{***}	0.073^{***}
	(0.002)	(0.004)	(0.006)	(0.008)	(0.009)
Year $\times$ Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓
Placebos: with randomized timing					
Process IDE	✓	✓	✓	✓	✓
Product IDE	✓	✓	✓	✓	✓
Observations	50,028	45,970	42,522	39,567	36,963
R^2	0.134	0.151	0.146	0.146	0.152

TABLE 14: Placebo 1. This table shows the association between $\overline{\text{IDE}}$ (PC_1, the composite displacement) and IDE^{tilt} (PC_2, the asymmetry component)—the first and second principal components of process and product IDE—and profit growth over the next five years (one to five years ahead), while controlling for placebo IDE measures constructed by randomizing the temporal ordering of each firm's technology embeddings within the firm before computing IDE. We 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 variable cost; current capital stock (Compustat ppegt, 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 double-clustered at the firm and year levels. The data are from 1997 to 2019. All firm variables are winsorized at the 1 percent level.

Firm profit growth	Forward Horizon (k years)				
$\log \Pi_{t+h} - \log \Pi_t$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A: IDE shifted by 1 year (IDE attached to year t is from year $t-1$)					
$\overline{\text{IDE}}$ (PC_1), shifted by 1 yr	-0.060^{***}	-0.105^{***}	-0.140^{***}	-0.162^{***}	-0.201^{***}
	(0.010)	(0.017)	(0.023)	(0.029)	(0.034)
IDE^{tilt} (PC_2), shifted by 1 yr	-0.000	-0.001	-0.003	-0.003	-0.004
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)
Firm innovation value	0.026^{***}	0.039^{***}	0.052^{***}	0.060^{***}	0.070^{***}
	(0.002)	(0.005)	(0.006)	(0.008)	(0.010)
Observations	44,204	40,789	37,918	35,426	33,206
R^2	0.134	0.148	0.145	0.141	0.149
Panel B: IDE shifted by 3 years (IDE attached to year t is from year $t-3$)					
$\overline{\text{IDE}}$ (PC_1), shifted by 3 yr	-0.026^{***}	-0.056^{***}	-0.071^{***}	-0.091^{***}	-0.100^{***}
	(0.009)	(0.016)	(0.023)	(0.028)	(0.034)
IDE^{tilt} (PC_2), shifted by 3 yr	-0.003^{*}	-0.003	-0.004	-0.004	-0.004
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)
Firm innovation value	0.022^{***}	0.037^{***}	0.051^{***}	0.062^{***}	0.071^{***}
	(0.003)	(0.005)	(0.007)	(0.009)	(0.011)
Observations	36,165	33,632	31,447	29,509	27,789
R^2	0.140	0.156	0.150	0.146	0.150
Panel C: IDE shifted by 5 years (IDE attached to year t is from year $t-5$)					
$\overline{\text{IDE}}$ (PC_1), shifted by 5 yr	-0.013	-0.033^{*}	-0.040^{*}	-0.046	-0.052
	(0.010)	(0.017)	(0.024)	(0.029)	(0.034)
IDE^{tilt} (PC_2), shifted by 5 yr	-0.000	-0.000	-0.001	0.002	-0.001
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)
Firm innovation value	0.028^{***}	0.047^{***}	0.065^{***}	0.078^{***}	0.089^{***}
	(0.003)	(0.006)	(0.009)	(0.012)	(0.014)
Observations	30,416	28,446	26,692	25,134	23,736
R^2	0.152	0.166	0.161	0.158	0.161

TABLE 15: Placebo 2. Profit growth on the baseline $\overline{\text{IDE}}$ (PC_1, the composite displacement) and IDE^{tilt} (PC_2, the asymmetry component) factors—the first and second principal components of process and product IDE—intentionally shifted in time. In each panel the IDE attached to a given firm-year is the value of IDE originally measured for that firm one, three, or five years earlier, depending on the panel. This is a placebo / persistence diagnostic: because IDE is highly autocorrelated, lagged IDE may continue to predict profit growth, but the magnitudes trace out how fast the predictive content decays. Controls, fixed effects, and clustering are identical to the main analysis: 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 ppegt, in logs), current employment (Compustat emp, in logs); and product-weighted sales growth. Fixed effects are sector-by-year plus industry. Standard errors are double-clustered at the firm and year level.

Firm profit growth	Forward Horizon (k years)				
$\log \Pi_{t+h} - \log \Pi_t$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$\overline{\text{IDE}}$ (PC_1), unweighted	-0.043^{***}	-0.068^{***}	-0.084^{***}	-0.101^{***}	-0.119^{***}
	(0.004)	(0.007)	(0.009)	(0.011)	(0.014)
IDE^{tilt} (PC_2), unweighted	-0.002	-0.001	0.002	-0.002	-0.001
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)
Firm innovation value	0.026^{***}	0.042^{***}	0.053^{***}	0.062^{***}	0.072^{***}
	(0.002)	(0.004)	(0.006)	(0.008)	(0.009)
Year × 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 16: This table shows the association between the unweighted IDE measures and profit growth over the next five years (one to five years ahead). $\overline{\text{IDE}}$ (PC_1) is the composite displacement and IDE^{tilt} (PC_2) 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 ppegt, 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 double-clustered at the firm and year levels. The data are from 1997 to 2019. All variables are winsorized at the 1 percent level.

Firm profit growth	Forward Horizon (k years)				
$\log \Pi_{t+h} - \log \Pi_t$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Panel A: $\overline{\text{IDE}}$ and IDE^{tilt}, high (2SD above mean)					
$\overline{\text{IDE}}$, high (PC_1)	-0.009^{***}	-0.023^{***}	-0.033^{***}	-0.038^{***}	-0.050^{***}
	(0.003)	(0.005)	(0.007)	(0.009)	(0.010)
IDE^{tilt}, high (PC_2)	-0.012^{***}	-0.016^{***}	-0.023^{***}	-0.023^{***}	-0.017^{**}
	(0.003)	(0.004)	(0.006)	(0.007)	(0.008)
Observations	50,028	45,970	42,522	39,567	36,963
R^2	0.132	0.149	0.144	0.143	0.149
Panel B: $\overline{\text{IDE}}$ and IDE^{tilt}, moderate (1SD; controlling for high)					
$\overline{\text{IDE}}$, moderate (PC_1)	-0.013^{***}	-0.021^{***}	-0.021^{**}	-0.029^{***}	-0.030^{**}
	(0.004)	(0.007)	(0.009)	(0.011)	(0.012)
IDE^{tilt}, moderate (PC_2)	-0.005^{*}	-0.002	-0.005	-0.008	-0.012
	(0.003)	(0.005)	(0.006)	(0.007)	(0.008)
$\overline{\text{IDE}}$, high (PC_1)	0.000	-0.009	-0.018^{**}	-0.019^{*}	-0.029^{**}
	(0.004)	(0.006)	(0.009)	(0.011)	(0.012)
IDE^{tilt}, high (PC_2)	-0.008^{***}	-0.014^{***}	-0.019^{***}	-0.017^{**}	-0.008
	(0.003)	(0.005)	(0.006)	(0.007)	(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)	-0.148^{***}	-0.234^{***}	-0.254^{***}	-0.304^{***}	-0.355^{***}
	(0.016)	(0.027)	(0.035)	(0.042)	(0.050)
IDE^{tilt} (PC_2)	-0.006^{***}	-0.007^{*}	-0.003	-0.010	-0.009
	(0.002)	(0.004)	(0.005)	(0.006)	(0.007)
$\overline{\text{IDE}}$, moderate (PC_1)	0.015^{***}	0.017^{**}	0.012	0.015	0.017
	(0.004)	(0.007)	(0.009)	(0.012)	(0.014)
IDE^{tilt}, moderate (PC_2)	0.004	0.013^{**}	0.008	0.010	0.014
	(0.003)	(0.005)	(0.007)	(0.008)	(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 17: This table reports the association between future profit growth and $\overline{\text{IDE}}$ (PC_1, the composite displacement) and IDE^{tilt} (PC_2, 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 double-clustered at the firm and year levels. The sample period is 1997 to 2019. All variables are winsorized at the 1 percent level.

Firm profit growth	Forward Horizon (k years)				
$\log \Pi_{t+h} - \log \Pi_t$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$\overline{\text{IDE}}$ (PC_1)	-0.102^{***}	-0.148^{***}	-0.180^{***}	-0.219^{***}	-0.240^{***}
	(0.015)	(0.025)	(0.034)	(0.041)	(0.050)
IDE^{tilt} (PC_2)	-0.000	0.003	0.008	0.006	0.007
	(0.003)	(0.005)	(0.006)	(0.008)	(0.009)
Firm innovation value	0.027^{***}	0.043^{***}	0.058^{***}	0.067^{***}	0.078^{***}
	(0.003)	(0.005)	(0.007)	(0.009)	(0.011)
Year × 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 18: This table shows the association between residual IDE measures and profit growth over the next five years (one to five years ahead). $\overline{\text{IDE}}$ (PC_1) is the composite displacement and IDE^{tilt} (PC_2) 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 ppegt, 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 double-clustered at the firm and year levels. The data are from 1997 to 2019. All variables are winsorized at the 1 percent level.

Appendix

- Section A describes the dataset construction and variable definitions. Section A.4, A.5, and A.6 contain prompt design details, sentence pairs for SemanticAxes, and example summaries. Section A.7 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.

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 Compustat's global dataset 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.

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 30 sectors of the Fama-French data library: 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; 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 2 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 67,050 firm-year observations on 8,702 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 30 industry categories defined by Fama-French.
- 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.

- Total (product) similarity (TS) 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 the sector.
- 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 $\text{IDE}^{\text{process}}_{f,t}$ and $\text{IDE}^{\text{product}}_{f,t}$ 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 $\texttt{at}_{f,t}$. 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 $k \in \{1,\ldots,5\}$ 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 calendar year, separately for each variable.

Winsorization before aggregation. Before forming asset-weighted industry means we winsorize firm-level $\text{IDE}^{\text{process}}_{f,t}$, $\text{IDE}^{\text{product}}_{f,t}$, the firm-level innovation value $A^{f}_{f,t}$, and Compustat `at` at the first and ninety-ninth percentiles on each tail of the full Compustat 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 $\text{IDE}^{\text{tilt}}_{i,t}$; 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 $\overline{A^f}_{i,t}$ as the asset-weighted industry mean of $A^f_{f,t}$ on the same sample, then percentile-rank it in $[0, 1]$ with ties averaged the same way. We include $\overline{A^f}_{i,t}$ 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 12 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 10-K prompts

A.4.1 Prompt to extract technology descriptions

Stage 1:

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. *Notes to Financial Statements*
6. *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.4.2 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.4.3 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.4.4 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.4.5 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.4.6 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.5 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.6 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.7 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}

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 A.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.

To keep the two channels of displacement visible—moderators may work asymmetrically across the process and product margins—we estimate the interactions in channel-specific form (process

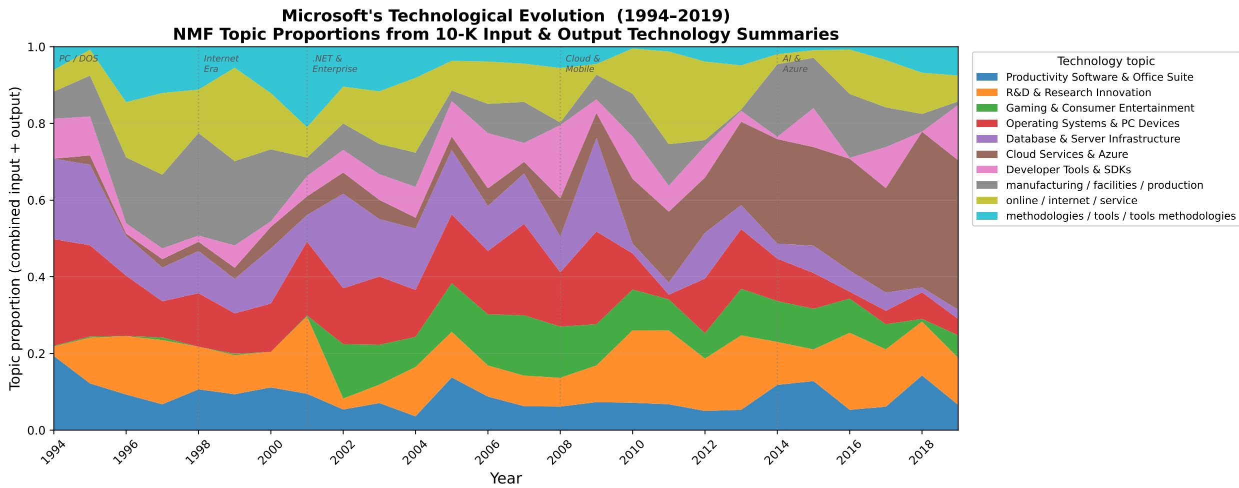

Figure A.1: This figure shows the evolution of technology topics mentioned in Microsoft's 10-K filings from 1997 to 2019.

and product IDE) rather than on the composite. For each horizon $k \in \{1, \ldots, 5\}$, we estimate

$$\begin{aligned}\log \Pi_{f,t+k} - \log \Pi_{f,t} = {} & \beta_1^k \, \mathrm{IDE}_{f,t}^{\text{process}} + \beta_2^k \, \mathrm{IDE}_{f,t}^{\text{product}} + \beta_3^k \, \mathrm{IDE}_{f,t}^{\text{process}} \, \mathrm{MkCap}_{f,t} + \beta_4^k \, \mathrm{IDE}_{f,t}^{\text{product}} \, \mathrm{MkCap}_{f,t} \\ & + \beta_5^k \, \mathrm{IDE}_{f,t}^{\text{process}} \, \mathrm{IntanCap}_{f,t} + \beta_6^k \, \mathrm{IDE}_{f,t}^{\text{product}} \, \mathrm{IntanCap}_{f,t} + \beta_7^k \, \mathrm{IDE}_{f,t}^{\text{process}} \, \mathrm{Profitability}_{f,t} \\ & + \beta_8^k \, \mathrm{IDE}_{f,t}^{\text{product}} \, \mathrm{Profitability}_{f,t} + \boldsymbol{\lambda}^{k\top} \mathbf{X}_{f,t} + \delta_{s \times t}^k + \gamma_g^k + \epsilon_{f,t}^k, \end{aligned} \tag{20}$$

where $\mathrm{MkCap}_{f,t}$ (market capitalization in billions) and $\mathrm{IntanCap}_{f,t}$ (intangible capital) enter as direct controls in $\mathbf{X}_{f,t}$ alongside the baseline controls in Equation (11), and $\mathrm{Profitability}_{f,t}$ is already in the baseline control set. Fixed effects, clustering, and standardization follow the headline specification.

Table A.1 reports the estimates. The direct effects of process and product IDE remain negative at every horizon. The interactions with intangible capital and profitability are individually small and largely insignificant. Market capitalization is the only characteristic that loads non-trivially on either interaction, and the loadings work in opposite directions across the two channels: larger firms appear less exposed to process displacement and more exposed to product displacement, with the channels offsetting so that the implied interaction with composite displacement is muted. We read this pattern as composition—larger firms tilt more toward product technologies—rather than a structural difference in displacement sensitivity, and we do not emphasize it in the body. The interaction with PMS in Section 5.1 remains the only firm-level moderator that produces an economically meaningful and consistent pattern across horizons.

D Model Appendix

This appendix develops a model of how other firms' innovations reduce the focal firm's profits through two channels: a product channel (product quality competition) and a process channel (technology-specific factor market tightening). We derive closed-form expressions for each channel and aggregate to the firm level. The firm-level results motivate the empirical IDE measures constructed in Section 2.3.

Firm profit growth	Forward Horizon (k years)				
$\log \Pi_{t+h} - \log \Pi_t$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
IDE process	-0.066^{**}	-0.096^{***}	-0.106^{***}	-0.132^{***}	-0.109^{*}
	(0.024)	(0.031)	(0.034)	(0.042)	(0.058)
IDE product	-0.016	-0.031	-0.044	-0.039	-0.093^{*}
	(0.020)	(0.028)	(0.032)	(0.038)	(0.052)
IDE process:mkcap	0.034	0.086^{*}	0.124^{*}	0.135^{**}	0.162^{**}
	(0.022)	(0.045)	(0.065)	(0.056)	(0.062)
IDE product:mkcap	-0.053^{**}	-0.103^{**}	-0.130^{*}	-0.144^{**}	-0.177^{**}
	(0.019)	(0.047)	(0.066)	(0.057)	(0.061)
IDE process:intangible capital	-0.005	-0.011	-0.021^{**}	-0.015	-0.021
	(0.007)	(0.010)	(0.009)	(0.011)	(0.013)
IDE product:intangible capital	0.010	0.016	0.024^{**}	0.021^{*}	0.030^{*}
	(0.007)	(0.010)	(0.009)	(0.011)	(0.014)
IDE process:profitability	0.033	0.071	0.077	0.067	0.015
	(0.035)	(0.056)	(0.065)	(0.081)	(0.105)
IDE product:profitability	-0.015	-0.034	-0.024	-0.023	0.043
	(0.034)	(0.053)	(0.062)	(0.075)	(0.095)
Year × Sector	✓	✓	✓	✓	✓
Industry	✓	✓	✓	✓	✓
Observations	33,394	30,956	28,885	27,135	25,489
R^2	0.128	0.144	0.134	0.126	0.135

TABLE A.1: Channel-specific (process/product) IDE and profit growth, with interactions between each channel and firm characteristics. The table reports estimates of Equation (20) for horizons $t+1$ through $t+5$. "IDE process" is $\text{IDE}^{\text{process}}$ and "IDE product" is $\text{IDE}^{\text{product}}$, each interacted with market capitalization (in billions), intangible capital, and current profitability. Controls include product-weighted sales growth, aggregate firm-level innovation value (deflated by assets), current profitability $(sale_{f,t} - cogs_{f,t})/at_{f,t}$, log current profit, current capital stock, intangible capital, market capitalization in billions, and current employment. The two IDE measures, the firm-level innovation value, and product-weighted sales growth are standardized to unit standard deviation. Fixed effects are sector-by-year plus 3-digit SIC industry. Standard errors are double-clustered by firm and year. Sample: 1997–2019.

D.1 Environment

There is a continuum of product markets indexed by m. In each market m there is a set of active firms $\mathcal{F}_m$. Firm f participates in a set of markets M_f. Within each market, firms compete under monopolistic competition.

Other firms' innovations affect the focal firm through two channels:

(i) **Product channel:** rivals increase product quality Q_{gm} (for $g \neq f$), shifting demand within market m away from firm f.

(ii) **Process channel:** other firms increase aggregate technology capital $K(j)$ in process j, tightening the process-j labor market and raising the focal firm's costs.

D.2 Demand

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

$$X_m = \left(\sum_{f \in \mathcal{F}_m} \left(Q_{fm} Y_{fm} \right)^{\frac{\theta_m - 1}{\theta_m}} \right)^{\frac{\theta_m}{\theta_m - 1}}, \qquad \theta_m > 1, \tag{21}$$

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

Define quality-adjusted quantity and price:

$$\tilde{Y}_{fm} \equiv Q_{fm} Y_{fm}, \qquad \tilde{p}_{fm} \equiv \frac{p_{fm}}{Q_{fm}}. \tag{22}$$

The dual CES price index is

$$P_m \equiv \left(\sum_{g \in \mathcal{F}_m} \tilde{p}_{gm}^{1-\theta_m} \right)^{\frac{1}{1-\theta_m}}. \tag{23}$$

Standard CES demand gives

$$\tilde{Y}_{fm} = X_m \left(\frac{\tilde{p}_{fm}}{P_m} \right)^{-\theta_m}. \tag{24}$$

Revenue and expenditure shares. Revenue is $R_{fm} \equiv p_{fm} Y_{fm} = \tilde{p}_{fm} \tilde{Y}_{fm}$. Nominal expenditure in market m is $E_m \equiv P_m X_m = \sum_{g \in \mathcal{F}_m} R_{gm}$. Under CES, the within-market expenditure share is

$$\sigma_{fm} \equiv \frac{R_{fm}}{E_m} = \frac{\tilde{p}_{fm}^{1-\theta_m}}{\sum_{g \in \mathcal{F}_m} \tilde{p}_{gm}^{1-\theta_m}}, \qquad \sum_{g \in \mathcal{F}_m} \sigma_{gm} = 1, \tag{25}$$

so that $R_{fm} = E_m \sigma_{fm}$.

Revenue response to price changes. Log-differentiating σ_{fm} from (25):

$$d \log \sigma_{fm} = (1 - \theta_m) \big(d \log \tilde{p}_{fm} - d \log P_m \big). \tag{26}$$

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

$$d \log R_{fm} = d \log E_m + (1 - \theta_m) \big(d \log \tilde{p}_{fm} - d \log P_m \big). \tag{27}$$

Revenue growth equals expenditure growth plus the change in the firm's expenditure share; the latter depends on the firm's price change relative to the market price index.

D.3 Production

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

$$Y_{fm} = \left(\sum_{j \in J} y_{fm}(j)^{\frac{\psi - 1}{\psi}} \right)^{\frac{\psi}{\psi - 1}}, \qquad \psi > 1, \tag{28}$$

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

Cobb–Douglas within each process.

$$y_{fm}(j) = k(f,j)^{\alpha}\, \ell_{fm}(j)^{1-\alpha}, \qquad \alpha \in (0,1), \tag{29}$$

where $k(f,j)$ is firm f's process-j capital (shared across products) and $\ell_{fm}(j)$ is process-j labor employed by firm f in market m.

Technology-specific labor markets. Labor is specific to each process j and clears globally:

$$\sum_{f} \sum_{m \in M_f} \ell_{fm}(j) = L(j), \qquad L(j) = 1 \text{ (normalization)}. \tag{30}$$

Let $w(j)$ denote the equilibrium wage in process j.

Aggregate technology capital.

$$K(j) \equiv \sum_{f} k(f,j). \tag{31}$$

Other firms' innovations in process j increase $K(j)$:

$$dK(j) = \mu(j)K(j)\, dN_{j,t}, \qquad \mu(j) > 0, \tag{32}$$

where $dN_{j,t}$ is an innovation arrival in process j.

D.4 Pricing and profits

Firms set prices as a constant markup over marginal cost (standard CES result):

$$p_{fm} = \frac{\theta_m}{\theta_m - 1} MC_{fm}, \qquad \tilde{p}_{fm} = \frac{\theta_m}{\theta_m - 1} \frac{MC_{fm}}{Q_{fm}}. \tag{33}$$

Profits are a constant share of revenue:

$$\Pi_{fm} = \frac{1}{\theta_m} R_{fm}, \qquad \Rightarrow \quad d \log \Pi_{fm} = d \log R_{fm}. \tag{34}$$

Since profits are proportional to revenue, we work with revenue throughout.

D.5 Product channel: product-quality displacement

We derive the effect of rivals' quality improvements on the focal firm's profits.

Setup. Fix a focal firm–product pair (f,m). Rivals' innovations raise their product qualities $\{Q_{gm}\}_{g \in \mathcal{F}_m \setminus \{f\}}$, holding all marginal costs fixed on impact. Under constant markup (33), a quality improvement by rival g leaves p_{gm} unchanged but lowers the quality-adjusted price:

$$d \log \tilde{p}_{gm} = d \log p_{gm} - d \log Q_{gm} = -d \log Q_{gm}. \tag{35}$$

The focal firm's quality is held fixed: $d \log Q_{fm} = 0$, so $d \log \tilde{p}_{fm} = 0$.

Price index response. Log-differentiating the CES price index (23):

$$d \log P_m = \sum_{g \in \mathcal{F}_m} \sigma_{gm} \, d \log \tilde{p}_{gm} = - \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} \, d \log Q_{gm}, \tag{36}$$

where the second equality uses $d \log \tilde{p}_{fm} = 0$.

Revenue response. Using (27) with $d \log \tilde{p}_{fm} = 0$:

$$d \log R_{fm} = d \log E_m + (\theta_m - 1) d \log P_m. \tag{37}$$

When the market price index falls (rivals improve quality), the focal firm's revenue declines through the expenditure share channel; the magnitude is amplified by the elasticity of substitution.

Fixed-expenditure benchmark. Under the benchmark $d \log E_m = 0$ (Cobb–Douglas upper tier or fixed expenditure), substituting (36):

$$d \log \Pi^{\text{out}}_{fm} = d \log R_{fm} = -(\theta_m - 1) \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} \, d \log Q_{gm}. \tag{38}$$

Displacement is stronger when demand is more elastic (θ_m large) and when quality growth is concentrated among large rivals (σ_{gm} large).

General case with expenditure reallocation. Allowing for an upper-tier CES across markets with elasticity $\nu > 0$ and market expenditure shares ϑ_m, the expenditure response to a price-index change is $d \log E_m = (1 - \nu)(1 - \vartheta_m) d \log P_m$. Substituting:

$$d \log \Pi^{\text{out}}_{fm} = - \Big[(\theta_m - 1) - (\nu - 1)(1 - \vartheta_m) \Big] \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} \, d \log Q_{gm}. \tag{39}$$

Setting $\nu = 1$ recovers the benchmark (38).

Firm-level aggregation. Total firm profits are $\Pi_f = \sum_{m \in M_f} \Pi_{fm}$. Define profit weights $\pi_{fm} \equiv \Pi_{fm} / \Pi_f$, so $\sum_{m \in M_f} \pi_{fm} = 1$. Then:

$$d \log \Pi^{\text{out}}_f = - \sum_{m \in M_f} \pi_{fm} \, (\theta_m - 1) \sum_{g \in \mathcal{F}_m \setminus \{f\}} \sigma_{gm} \, d \log Q_{gm}. \tag{40}$$

This is the comprehensive product-channel equation: the firm's profit decline is a profit-weighted sum, across all its product markets, of the expenditure-share-weighted quality improvements by rivals in each market.

D.6 Process channel: factor-market displacement

We derive the effect of an increase in aggregate technology capital $K(j)$—driven by other firms' innovations, holding the focal firm's own $k(f, j)$ fixed—on the focal firm's profits. Product qualities $\{Q_{gm}\}$ are held fixed throughout.

D.6.1 Cost minimization and cost shares

For a given output target Y_{fm}, the firm chooses labor allocations $\{\ell_{fm}(j)\}_{j\in J}$ to minimize total cost $\sum_j w(j)\ell_{fm}(j)$ subject to the production constraints (28)–(29). Substituting the Cobb–Douglas into the CES:

$$Y_{fm} = \left(\sum_{j\in J} k(f,j)^{\frac{\alpha(\psi-1)}{\psi}}\, \ell_{fm}(j)^{\frac{(1-\alpha)(\psi-1)}{\psi}}\right)^{\frac{\psi}{\psi-1}}. \tag{41}$$

Since the production function is homogeneous of degree $(1-\alpha)$ in labor (by Euler's theorem applied to the CES of Cobb–Douglas sub-functions with fixed k), total cost satisfies

$$TC_{fm} = (1-\alpha)\, MC_{fm}\, Y_{fm}. \tag{42}$$

Total cost is proportional to output times marginal cost, with the factor $(1-\alpha)$ reflecting the labor share in production.

Cost shares. Define the cost share of technology j for firm f in market m:

$$\phi_{fm}(j) \equiv \frac{w(j)\,\ell_{fm}(j)}{TC_{fm}}. \tag{43}$$

From the first-order conditions of the CES cost minimization, one can show that

$$\phi_{fm}(j) = \frac{y_{fm}(j)^{(\psi-1)/\psi}}{\sum_{h\in J} y_{fm}(h)^{(\psi-1)/\psi}} = \frac{y_{fm}(j)^{(\psi-1)/\psi}}{Y_{fm}^{(\psi-1)/\psi}}. \tag{44}$$

This is the standard CES result: the cost share of process j equals its contribution to the CES aggregator relative to total output. Crucially, the cost shares do not depend on the product market m—they are identical across all products of the same firm, because $k(f,j)$ is shared. We therefore write $\phi_f(j) \equiv \phi_{fm}(j)$ for all $m \in M_f$.

Marginal cost elasticity (envelope theorem). By the envelope theorem, the change in total cost holding output fixed equals $\sum_j \ell_{fm}(j)\,dw(j)$. Since $TC_{fm} = (1-\alpha)MC_{fm}Y_{fm}$ and Y_{fm} is held fixed:

$$d\log MC_{fm} = \sum_{j\in J} \phi_f(j)\, d\log w(j). \tag{45}$$

The marginal cost elasticity to wages is a cost-share-weighted sum.

D.6.2 Technology exposure shares

The cost shares $\phi_f(j)$ depend on the equilibrium labor allocations, which in turn depend on wages and technology capital. We now express them in terms of the primitive objects $k(f,j)$ and the equilibrium wages $w(j)$.

From the CES first-order conditions, the optimal technology-j output satisfies

$$y_{fm}(j) \propto \left(\frac{w(j)}{k(f,j)^{\alpha/(1-\alpha)}}\right)^{-\tilde{\psi}(1-\alpha)} \quad \text{where} \quad \tilde{\psi} \equiv \frac{\psi}{1+\alpha(\psi-1)}. \tag{46}$$

The parameter $\tilde{\psi}$ is the effective elasticity of substitution across processes, adjusted for the decreasing returns to labor within each process. Substituting into (44):

$$\phi_f(j) = \frac{\left[k(f,j)^{\alpha/(1-\alpha)}/w(j)\right]^{(\tilde{\psi}-1)(1-\alpha)}}{\sum_{h\in J}\left[k(f,h)^{\alpha/(1-\alpha)}/w(h)\right]^{(\tilde{\psi}-1)(1-\alpha)}}. \tag{47}$$

The cost share of technology j is higher when the firm has more capital in process j relative to that process's wage.

Equilibrium wages. From the technology-specific labor market clearing condition (30), the equilibrium wage in process j satisfies

$$w(j) \propto K(j)^{\alpha} \cdot K(j)^{-(1-\alpha)/\tilde{\psi}} \cdot (\text{terms common to all } j), \tag{48}$$

where the proportionality reflects the dependence of wages on aggregate demand conditions and the common Lagrange multiplier. The key property is that an increase in $K(j)$ raises $w(j)$: more technology capital in process j raises productivity in that process, which increases labor demand and bids up the process-specific wage.

Substituting the equilibrium wage structure into (47), the cost shares simplify. Define

$$S_{f,h} \equiv \left[k(f,h)^{\alpha}\left(\frac{k(f,h)}{K(h)}\right)^{1-\alpha}\right]^{\frac{\psi-1}{\psi}}, \qquad S_f \equiv \sum_{h\in J} S_{f,h}, \qquad s_{f,h} \equiv \frac{S_{f,h}}{S_f}. \tag{49}$$

Then $\phi_f(j) = s_{f,j}$: the cost share of technology j equals the technology exposure share $s_{f,j}$. Intuitively, $s_{f,j}$ measures the fraction of firm f's effective production capability drawn from process j, accounting for both its own capital $k(f,j)$ and its share of the aggregate capital pool $k(f,j)/K(j)$.

Competitor and economy-wide exposure. Define the expenditure-weighted average exposure in market m:

$$\bar{s}_{m,j} \equiv \sum_{g\in\mathcal{F}_m} \sigma_{gm}\, s_{g,j}, \tag{50}$$

and the economy-wide technology weight:

$$\omega_j \equiv \frac{K(j)^{-\frac{\alpha(1-\psi)}{\psi}}}{\sum_{h\in J} K(h)^{-\frac{\alpha(1-\psi)}{\psi}}}, \qquad \sum_{j\in J}\omega_j = 1. \tag{51}$$

Processes with greater aggregate capability receive higher weight; the weights sum to one across all processes.

D.6.3 Wage response to an increase in $K(j)$

An increase in $K(j)$ affects all wages through the GE labor market system. Log-differentiating the labor market clearing conditions and using the equilibrium structure yields:

Own-process effect. An increase in $K(j)$ directly raises the marginal product of labor in process j, increasing $w(j)$:

$$\frac{\partial \log w(j)}{\partial \log K(j)} = \alpha + (1-\alpha)\left(1-\omega_j\right)\frac{\psi-1}{\psi}. \tag{52}$$

The own-process wage rises through both a direct productivity effect (α term) and a labor reallocation effect (the remaining term).

Cross-process effect. For $h \neq j$, the increase in $K(j)$ draws labor toward process j (through the CES substitution), reducing labor supply in other processes:

$$\frac{\partial \log w(h)}{\partial \log K(j)} = -(1-\alpha)\,\omega_j\,\frac{\psi-1}{\psi}. \tag{53}$$

The cross effect is negative (wages in other processes fall) and proportional to ω_j, the economy-wide weight of process j. In the limit $\psi \to \infty$ (perfect substitution across technologies), the cross effect vanishes.

D.6.4 Product-level profit elasticity

Combining the marginal cost response (45) with the wage responses (52)–(53), the focal firm's marginal cost response to an increase in $K(j)$ is:

$$\frac{\partial \log MC_{fm}}{\partial \log K(j)} = s_{f,j}\,\frac{\partial \log w(j)}{\partial \log K(j)} + \sum_{h\neq j} s_{f,h}\,\frac{\partial \log w(h)}{\partial \log K(j)} = \frac{\psi-1}{\psi}(1-\alpha)\,s_{f,j} + \alpha\,s_{f,j} - (1-\alpha)\frac{\psi-1}{\psi}\,\omega_j\sum_h s_{f,h}. \tag{54}$$

Simplifying using $\sum_h s_{f,h} = 1$:

$$\frac{\partial \log MC_{fm}}{\partial \log K(j)} = \frac{\psi-1}{\psi}(1-\alpha)\,s_{f,j} + \frac{\psi-1}{\psi}\alpha\,\omega_j. \tag{55}$$

This is the *partial-equilibrium* cost response: the focal firm's marginal cost rises in proportion to its exposure $s_{f,j}$ (direct effect) and the economy-wide weight ω_j (GE effect). Every competitor g in market m faces an analogous cost increase with $s_{g,j}$ in place of $s_{f,j}$.

Within-market price response. From constant markup, $d\log \tilde{p}_{gm} = d\log MC_{gm}$ for all $g \in \mathcal{F}_m$ (holding qualities fixed). The within-market CES price index responds as:

$$d\log P_m = \sum_{g\in\mathcal{F}_m} \sigma_{gm}\,d\log MC_{gm}. \tag{56}$$

Substituting (55) for each firm:

$$\frac{\partial \log P_m}{\partial \log K(j)} = \frac{\psi-1}{\psi}(1-\alpha)\,\bar{s}_{m,j} + \frac{\psi-1}{\psi}\alpha\,\omega_j. \tag{57}$$

The market price index rises with the average exposure of firms in the market ($\bar{s}_{m,j}$) and the economy-wide weight (ω_j).

Revenue and profit response. From (27) with $d \log E_m = 0$ (fixed expenditure):

$$\frac{\partial \log R_{fm}}{\partial \log K(j)} = (1-\theta_m)\left(\frac{\partial \log MC_{fm}}{\partial \log K(j)} - \frac{\partial \log P_m}{\partial \log K(j)}\right). \tag{58}$$

Substituting (55) and (57):

$$\frac{\partial \log R_{fm}}{\partial \log K(j)} = -(1-\theta_m)\cdot\frac{\psi-1}{\psi}(1-\alpha)\left(s_{f,j}-\bar{s}_{m,j}\right) - \frac{\partial \log MC_{fm}}{\partial \log K(j)}, \tag{59}$$

but it is cleaner to substitute directly. From (27) and (34):

$$\begin{aligned}
\frac{\partial \log \Pi_{fm}}{\partial \log K(j)} &= (1-\theta_m)\left(\frac{\partial \log MC_{fm}}{\partial \log K(j)} - \frac{\partial \log P_m}{\partial \log K(j)}\right) \\
&= -\frac{\psi-1}{\psi}(1-\alpha)\left\{\left[1+\frac{(\theta_m-\psi)(\theta_m-1)}{\theta_m}\right]s_{f,j} - \frac{(\theta_m-\psi)(\theta_m-1)}{\theta_m}\,\bar{s}_{m,j}\right\} \\
&\quad -\frac{\psi-1}{\psi}\alpha\,\omega_j.
\end{aligned} \tag{60}$$

Interpretation of the three terms. The first line gives the closed-form process-channel elasticity, which decomposes into three terms:

- **Direct cost effect** (proportional to $s_{f,j}$): the firm's marginal cost rises because its own production uses technology j.
- **Within-market reallocation** (proportional to $\bar{s}_{m,j}$): competitors in market m also face higher costs, partially offsetting the focal firm's disadvantage. The net effect depends on the firm's exposure *relative to* the market average: firms with $s_{f,j} > \bar{s}_{m,j}$ are hurt more.
- **Economy-wide GE effect** (proportional to ω_j): cross-technology substitution in factor markets raises all firms' costs proportionally.

Verification. The expression (60) is negative for all firms when $\psi > 1$ and $\theta_m > 1$: an increase in $K(j)$ driven by other firms always reduces the focal firm's profits. The magnitude is larger when the firm is more exposed to process j (higher $s_{f,j}$), when technologies are more substitutable (higher ψ), and when the innovating process is larger in the economy (higher ω_j).

D.6.5 Firm-level aggregation

Aggregation across products. Using profit weights $\pi_{fm} \equiv \Pi_{fm}/\Pi_f$:

$$\frac{\partial \log \Pi_f}{\partial \log K(j)} = \sum_{m\in M_f} \pi_{fm}\frac{\partial \log \Pi_{fm}}{\partial \log K(j)}. \tag{61}$$

Aggregation across processes. Using the innovation arrival process (32), the total process-channel effect on firm f's profits is:

$$d\log \Pi_f^{\text{process}} = \sum_{j\in J}\frac{\partial \log \Pi_f}{\partial \log K(j)}\,\mu(j)\,dN_{j,t}. \tag{62}$$

This is the comprehensive process-channel equation: the firm's profit decline is a sum across all processes of its profit elasticity to each process's capability, weighted by the innovation arrival rate and step size.

Equivalently, substituting (61) and (60):

$$d \log \Pi_f^{\text{process}} = \sum_{j \in J} \left[\sum_{m \in M_f} \pi_{fm} \frac{\partial \log \Pi_{fm}}{\partial \log K(j)} \right] \mu(j) \, dN_{j,t}, \tag{63}$$

which makes explicit the double aggregation over products and technologies.