Growth is Getting Harder to Find, Not Ideas*

Teresa C. Fort[†] Nathan Goldschlag[‡] Jack Liang[§] Peter K. Schott[¶]
Nikolas Zolas[‖]

March 26, 2026

Abstract

Relatively flat US output growth versus rising numbers of US researchers is often interpreted as evidence that ideas are getting harder to find. We build a new 45-year panel tracking the universe of US firms' patenting to investigate the micro underpinnings of this claim, separately examining the relationships between research inputs and ideas (patents) versus ideas and growth. We find that average patents per R&D input are increasing, the elasticity of patents to R&D inputs is flat or rising, and there is no systematic evidence of a secular decline in patenting after controlling for research inputs. We then document a positive, significant, and fairly steady relationship between firms' growth in ideas (patents) and labor productivity. Average firm growth after controlling for idea growth, however, declines. Together, these results suggest that innovative efforts play a key role in sustaining growth that has not diminished over the last four decades.

*We thank Cheryl Grim, Samuel Kortum, Bitsy Perlman, Michael Peters, Thomas Sampson, and numerous conference and seminar participants for helpful comments and suggestions. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the US Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7083300: CBDRB-FY25-CES002-001 and CBDRB-FY25-CED006-012). Part of the paper was written while Fort, Goldschlag, and Zolas were US Census Bureau employees.

[†]Tuck School at Dartmouth, US Census Bureau, NBER, and CEPR.
[‡]US Census Bureau.
[§]Temple University.
[¶]Yale School of Management, NBER, CEPR, CESifo.
[‖]US Department of State.

1 Introduction

A central question for understanding long-run growth is whether and how the existing stock of knowledge affects future growth. In his Nobel-prize-winning work, Romer (1990) shows that firms' profit-seeking R&D investments can affect the economy's long-run growth rate, even when the population is constant. This "endogenous growth" arises under the assumption that the number of ideas researchers discover each year rises linearly with the stock of existing ideas. If, instead, a larger stock of ideas makes it increasingly difficult to find new ones, firms will reallocate workers towards production and the economy's growth rate will stall. Romer acknowledges this crowding-out possibility and justifies his production function for knowledge noting that "...there is no evidence from recent history to support the belief that opportunities for research are diminishing."[1]

While Romer (1990) predicts higher output growth from greater R&D efforts, the US economy has maintained an average two percent real growth rate over the last 200 years despite substantial increases in its R&D spending. These patterns have led some researchers to conclude that ideas are getting harder to find (Jones, 1995; Kortum, 1997). Most recently, Bloom et al. (2020) infer declining researcher productivity from the divergence between rising R&D efforts versus flat US productivity growth for various samples of industries and firms. These trends can be rationalized by models in which researchers' marginal products exhibit decreasing returns in the aggregate stock of ideas. In those models, there is no role for innovation policy to affect long-run growth, since improvements in living standards depend solely on population growth.

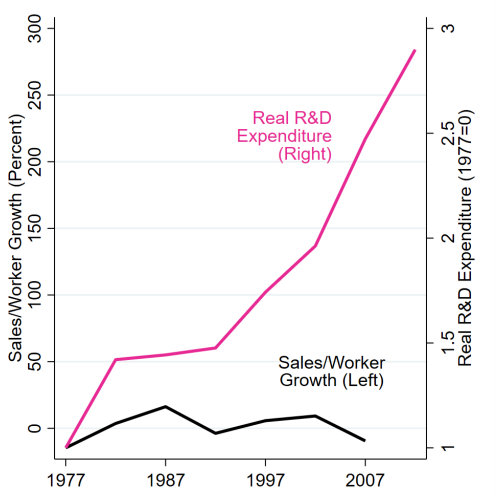
Although constant output growth coupled with rising research efforts is consistent with a growing stock of knowledge making it harder to innovate, existing papers do not provide direct evidence of this mechanism. Moreover, the inference that ideas are getting harder to find depends upon the simplifying assumption, common in this literature, that idea growth maps one-to-one with output growth. This assumption neglects the possibility that the relationship between ideas and growth may not be linear or constant over time. It also leaves no role for other factors, such as changing market structure or regulatory environments, to affect growth.

In this paper, we analyze whether ideas are getting harder to find by studying these two relationships separately: that between research inputs and ideas, versus that between ideas and growth. To pinpoint the underlying mechanisms at play in macro-models, we construct a novel firm-level panel from 1977 to 2021 with a rich set of R&D inputs using the universe of US private-sector employer businesses in the Census Bureau's Longitudinal Business Database (LBD). Figure 1 depicts R&D inputs and aggregate growth in our data, and delivers the same message highlighted in Bloom et al. (2020): US sales per worker since 1977 has remained relatively flat, while R&D inputs have risen threefold. To assess whether the divergence in Figure 1 arises from declining researcher efficiency, we merge the Census microdata to US patent grants, which we take as a key manifestation of ideas.

Our first contribution is to provide several new facts on US patenting. The first of these highlights the importance of studying the universe of US firms and quality-adjusted patent measures when analyzing research productivity. Early models with diminishing returns to aggregate knowledge stocks

[1] Romer (1990), page S84. Related papers with endogenous growth yield similar predictions (Grossman and Helpman, 1991; Aghion and Howitt, 1992).

Figure 1: Aggregated Micro Data on Growth and Research Inputs



Source: PV, LBD, EC, RADS, and authors' calculations.
Notes: Figure replicates Bloom et al. (2020) Figure 1 using aggregated US Census Bureau micro data analyzed in this paper. Left axis reports growth in aggregate sales per worker across firms in the LBD by semi-decade, where the point for 1977 us the average growth for 1977 to 1981. Right axis reports aggregate real R&D expenditure across firms in the LBD, by semi-decade, indexed to 1 for the 1977 to 1981 semi-decade.

were motivated by studies documenting declining patents per researcher over time (Evenson, 1984; Kortum, 1993; Lanjouw and Schankerman, 2004). While we also observe this pattern among the subset of publicly traded companies in Compustat, we find that the ratio of patents per real R&D dollar *rises* from 1977 to 2017 when considering *all* firms. We report similar increases in such research productivity using measures of patenting that account for quality: counts of patents' external citations (Akcigit and Kerr, 2018) and counts of "breakthrough" patents (Kelly et al., 2021).[2] Moreover, research productivity using these quality-adjusted measures also rises among the subset of Compustat firms, albeit less sharply. Our data thus reveal rising research productivity among the universe of all US firms, as well as among the subset of publicly traded firms when using quality-adjusted patents. These results do not support a first way of defining "ideas are getting harder to find," *i.e.,* that the average number of patents per R&D dollar has fallen since the 1970s.

We also provide two new facts on changes in the evolution of US patenting across firms' sectors, ages, and sizes. Although manufacturing firms are responsible for the majority of patenting in the 1970s and 1980s, they account for less than a third in the 2010s, with their share of breakthrough patents falling from 73 to 12 percent. Patenting activity is instead increasingly accounted for by firms in the Information, Management, and Professional, Scientific, and Technical Services sectors, which combined account for 45 and 66 percent of citations and breakthroughs, respectively, by the 2010s.

[2]We count external (rather than self) citations within five years of each patent's grant date. Breakthrough patents are those whose ratio of text dissimilarity to past patents versus similarity to future patents is in the top five percentile.

Patenting also changes markedly across the firm size and age distributions. "Mega-firms" with more than 10 thousand workers account for 58 percent of patent grants and external citations in the 1970s, but only 48 and 49 percent, respectively, by the 2010s. Their share of breakthroughs falls considerably during the internet-boom era, but finishes the sample period roughly where it started, at 67 percent. The changing composition of patenting firms coincides with a decrease in patent coverage by publicly traded firms, from 63 percent in the late 1970s to 55 percent by 2017. These trends indicate that past work focused solely on the US manufacturing sector or Compustat firms misses a growing share of US innovative outputs over time.

Our second contribution is to estimate potentially time-varying elasticities of patents to R&D inputs. To account for the changing composition of patenters highlighted above, we focus on within-firm patent elasticities. In contrast to the notion that the marginal product of R&D declines over time, we find rising elasticities of patent grants, citations, and breakthroughs to R&D expenditure. For patent grants, elasticities increase from 0.26 to 0.48 over the 1977 to 2016 period. Although these estimates are not causal, our use of firm fixed effects ensures that we capture a rise in subsequent patents when a particular firm increases its R&D spending. And, while the level of these estimates may be biased upwards if firm-by-year shocks raise both R&D and patenting, the size of such shocks would need to rise systemically over time to match the elevating elasticities we document – a pattern which itself would seem to contradict the notion that ideas are getting harder to find.

Census R&D surveys disproportionately sample a subset of large manufacturing firms, which our descriptive evidence shows is less relevant for patenting over time. To ensure that our results are not driven by selection into R&D surveys, we construct new measures of firms' R&D inputs using their payroll in innovation-related establishments, such as R&D labs (as a narrow measure) and Management, Professional, Scientific, and Technical Services establishments (as a broader measure).[3] Whereas patent coverage by firms in the R&D surveys falls from 65 to 59 percent over our period, firms with Management and Professional Services establishments account for about 77 percent of patent grants throughout. We also consider firms' total payroll as a proxy for firms' R&D inputs; although broad, this measure allows us to estimate relationships among all US patenting firms. Estimated patent elasticities across all three patent activities and all four measures of research inputs are generally flat or rising over time. These results do not support a second way of defining "ideas are getting harder to find," *i.e.,* that the marginal R&D dollar is less effective at generating an idea over time.

The process of patenting might change for various reasons over our decades-long sample, *e.g.*, in terms of the quality of the ideas firms seek to protect, or how the US Patent and Trademark Office (USPTO) grants approvals. We include semi-decade fixed effects in our specifications to ensure that our patenting elasticities are not contaminated by institutional or macroeconomic shocks that affect all firms equally. Trends in these fixed effects may also indicate that a rising stock of ideas makes finding new ones more difficult. In particular, a secular flattening or decline in these fixed effects indicates fewer patents per firm over a period that coincides with a rising aggregate stock of patents.

[3]R&D labs are classified in NAICS 5417. Management is NAICS 55, Professional, Scientific, and Technical Services is NAICS 54, which includes R&D labs as one industry.

While the data reveal such declines for two of our six specifications using patent grants, the time fixed effects for citation-weighted grants are flat or rising across all measures of research inputs. Those for breakthrough patents are flat or rising for the first half of our sample period and decline thereafter. These trends do not provide systematic support for a third way of defining "ideas are getting harder to find," *i.e.*, that idea creation has decreased universally across all patenting firms independent of their research efforts over the last four decades.

Our third contribution is to estimate a potentially time-varying relationship between idea and output growth for all US patenting firms. We regress five-year growth rates in a firm's sales on the growth in its patent stock and employment. Given the falling relevance of manufacturing in US patenting, we include all sectors in this analysis and proxy for any changes in firms' capital stocks using size and age bins. We also consider specifications using sales per worker as the dependent variable. We find a positive and statistically significant relationship for both outcomes in most periods. Most notably, there is no weakening in either association over time. Indeed, the relationship between patent growth and sales per worker growth is strongest in the later years, and particularly so for breakthrough patents. These results do not support a fourth way of defining "ideas are getting harder to find," *i.e.,* a weakening of the relationship of between ideas and growth. They also address the concern that idea quality has deteriorated over time, since such a deterioration should manifest as a weakening association between ideas and growth.

We use the sales growth regression estimates to decompose the predicted effects on sales growth of changes in firms' ideas, employment, and size and age bins, along with the semi-decade fixed effects. While ideas predict fairly consistent growth throughout our sample period, the other factors tend to predict *lower* overall growth in later periods. Notably, the semi-decade fixed effects and employment coefficients both predict relatively low aggregate growth rates, particularly during the 2000s. Indeed, firms' own ideas predict positive sales growth in the last decade, even as other factors drag growth down.

Together, our results indicate that researcher productivity is rising, the marginal R&D dollar is at least as effective in the 2010 decade as it was in the 1970s, there is no systematic secular decline in firms' ability to find ideas after controlling for their R&D input use, and that, if anything, the link between ideas and productivity growth is rising. These patterns are hard to square with the claim that ideas are getting harder to find. On the other hand, we find that ideas play a fairly small role in predicting aggregate firm growth, with other factors dominating at different periods. Since the falling semi-decade fixed effects we estimate in the growth regressions capture changes common to all firms, one possibility is that the impact of spillovers on growth has fallen for all firms, despite a robust relationship between a firm's own ideas and its growth.

This paper contributes to three strands of the literature. First, we add to a large literature that estimates the elasticity of patents to R&D expenditure, with early work summarized by Griliches (1990) and further explored by Blundell et al. (2002). Most related to our analysis on changing elasticities, Hausman et al. (1984) estimate declining patent elasticities for a sample of 128 firms from 1968 to 1974. More recent papers exploit exogenous variation in R&D funding to assess its impact on firm innovation and patenting (Howell, 2017; Meyers and Lanahan, 2022). Our contribution is to

estimate flat or rising within-firm elasticities of patent grants, external citations, and breakthroughs to comprehensive measures of R&D inputs for all firms in the United States over the last four decades.[4]

We also expand on a relatively small set of papers studying the relationship between firm-level patenting and growth. While a long line of research in industrial organization estimates the effect of R&D expenditure on firm productivity (Griliches, 1979; Hall et al., 1998; Aw et al., 2011; Peters et al., 2017; Chen and Xu, 2023; Ando et al., 2025), there is less evidence on whether and how a firm's patents relate to subsequent growth. Balasubramanian and Sivadasan (2011) link manufacturing firms' patenting to increases in their size and productivity. Kline et al. (2019) show that a firm's first patent grant leads to higher profits. Patents grants are also associated with increased stock market value (Hall et al., 2005), though with substantial variation in those returns, with higher returns correlating with future growth (Kogan et al., 2017). Akcigit and Kerr (2018) document a positive relationship between the log of patents and employment growth for US firms over the 1982 to 1997 period. We document a positive and significant relationship between a firm's patents and its labor productivity growth that strengthens over the last four decades for the universe of US firms.

Finally, we build on a growing body of empirical work that studies whether flat (or falling) US productivity growth coupled with rising research efforts may be due to negative crowding-out effects of a growing knowledge stock.[5] Several influential papers document the rising divergence between productivity growth and research inputs and rationalize it in a framework in which a larger stock of knowledge makes it harder to find new ones (Jones, 1995; Kortum, 1997; Bloom et al., 2020).[6] Jones (2009) provides microevidence that it is harder to innovate as a particular field matures, which he models as decreasing output growth from innovative effort. Most recently, Ekerdt and Wu (2024) show that declining aggregate researcher productivity occurs when heterogeneous workers increasingly select into innovation rather than production, but that such declines are transient. Our contribution is to study all US firms and rule out four different potential microchannels through which ideas getting harder to find could manifest in the data.

This paper also facilitates future empirical work by providing a new crosswalk between the LBD and the USPTO patent data that is available to all researchers with approved projects through the Federal Statistical Research Data Centers (FSRDCs). Our bridge provides the longest period of matched data, spanning 1976 to 2021, matches 92 percent of US-based patent-assignee records, and ensures longitudinal consistency in the matching process. To our knowledge, we are the first to develop and share such longitudinal matching techniques, which can be generalized to improve merges of other external datasets to the Census Bureau's micro data. The crosswalk is also used to produce the Census Bureau's public-use Business Dynamics Statistics of Patenting Firms tables, which are publicly available tabulations of US job creation and destruction by firms' patent status.

The remainder of this paper is structured as follows. Section 2 and 3 provide background for our empirical specifications and describe the data used to estimate them. Section 4 summarizes patenting activity by US firms from 1977 to 2017, comparing results for all US firms in our LBD panel to

[4]Our rising estimates are indirectly suggestive that idea growth is not linear as suggested by Philippon (2022).

[5]Books on this topic include (Cowen, 2011; Brynjolfsson and McAfee, 2014; Gordon, 2017).

[6]Some recent papers propose that R&D labor and R&D capital are complementary inputs into the idea production function, with declines in the latter making ideas harder to find (Growiec et al., 2022; Ekerdt, 2024).

an analogously constructed Compustat panel. Sections 5 and 6 describe our estimation of patent elasticities and the relationship between patenting and firm growth. Section 7 concludes.

2 From Macro Growth Theory to Firm Regressions

Romer (1990) presents a model in which firms' profit-maximizing decisions to innovate drive long-run growth. In his setting, the aggregate knowledge stock, or number of 'ideas' (A), affects productivity in final-good production and evolves according to

$$\dot{A} = \delta H_A A, \tag{1}$$

where $\dot{A}$ is the flow of new ideas, H_A is total human capital (or researchers) devoted to innovation, and δ is a productivity parameter. This formulation is built on the premise that an idea is nonrival and only partially excludable: every new idea provides a potential building block for subsequent ideas. As a result, the flow of future ideas is increasing in the aggregate knowledge stock.

Final output is produced by a representative firm using the following Cobb-Douglas technology

$$Y(H_Y, L, x) = H_Y^{\alpha} L^{\phi} \sum_{i=1}^{A} x_i^{1-\alpha-\phi}, \tag{2}$$

in which H_Y is human capital devoted to final-good production, L is labor services, and x_i is a durable capital-good input. Each input, x_i requires an idea i (or blueprint) to produce, with monopolistically competitive input producers converting ω units of foregone consumption into one unit of inputs. Although there is an infinite potential number of ideas, only those that have been discovered – denoted by A in equation (1) – can be used to produce. Crucially, and in contrast to prior growth models, there is no decreasing returns to such capital as new varieties of x_i are added. The aggregate knowledge stock thus determines the number of capital-good inputs in the economy, which in turn affects output. In particular, equation (2) can be rewritten as

$$Y(H_Y, L, x) = (H_Y A)^{\alpha} (LA)^{\phi} (K)^{1-\alpha-\phi} \omega^{\alpha+\phi-1}, \tag{3}$$

where K denotes physical capital of durable inputs and is thus the physical embodiment of the stock of ideas.[7] Differentiating equation (3) with respect to time, holding both types of labor fixed, and assuming that ω is time-invariant, the steady-state growth rate of output is determined simply by the growth rate of ideas, which in turn is determined solely by researcher productivity and the amount of labor dedicated to human capital in knowledge production, i.e.,

$$\frac{\dot{Y}}{Y} = \frac{\dot{A}}{A} = \delta H_A. \tag{4}$$

The model thus predicts a higher growth rate when the number of innovation workers rises.

[7] An intermediate step between equations (2) and (3) is $Y(H_Y, L, x) = A H_Y^{\alpha} L^{\phi} \bar{x}^{1-\alpha-\phi}$, from which it is clear that equation (4) follows.

As discussed in the introduction, subsequent macro models have focused on the fact that this prediction is severely counterfactual, as the number (and share) of researchers in the US economy has increased substantially, while aggregate growth has remained fairly constant. Jones (1995) proposes a 'semi-endogenous' growth model with diminishing returns to new ideas

$$\dot{A} = \delta H_A A^{1-\beta} \tag{5}$$

as a way to reconcile Romer (1990) with this macro fact. In this formulation, $\beta > 0$ implies that the growth rate of ideas ($\dot{A}/A$) will decline over time as the existing stock of knowledge expands.[8] More recently, Bloom et al. (2020) define *Researcher Productivity* $\coloneqq \frac{\dot{A}_t/A_t}{H_{At}} = \delta A_t^{-\beta}$ and infer the size of β under the null hypothesis that this ratio (and δ) remain constant. They measure $\dot{A}_t/A_t$ as US TFP growth (as well as using various sector measures of productivity growth, such as greater crop yields or higher life expectancy) and H_A as increases in researchers in the US economy or a particular sector. To the extent that their measure of $\frac{\dot{A}_t/A_t}{H_{At}}$ falls over time, the authors infer declining researcher productivity, which they attribute to $\beta > 0$.[9] Indeed the formulation in equation (5) is elegant in that they can always 'match' the data perfectly by adjusting β for different time periods or sectors. On the other hand, this approach entails a time-period and sector-specific β, which does not quite align with the formulation in equation (5).

Although a time-varying β in equation (5) is one way to reconcile equations (1) and (3) with aggregate trends, at least three alternative explanations are worthy of attention. First, researcher productivity might change over time for a variety of reasons, including policy. To the best of our knowledge, the empirical research in the growth literature does not tie the inferred decline in researcher productivity directly to increases in the *aggregate* stock of ideas (A).[10] A second possibility is that the relationship between idea growth and output growth is not one-to-one or constant over time. Indeed, Romer (1990) recognizes that equation (2), in which all durable inputs have additively separable effects on final-good output, may not hold in practice and notes, "An investigation of complementarity as well as a of mixtures of types of substitutability is left for future work" (p. S81). Finally, growth depends not only on ideas, but also on other factors, such as changing regulation (Djankov et al., 2002), market structure (Aghion et al., 2005), climate (Dell et al., 2012), and institutions (Acemoglu et al., 2001), among others. It is possible that idea growth is expanding, and ideas continue to generate productivity improvements, but other factors work against these forces.

To investigate the empirical support for the first two channels – a decline in researchers' ability to generate ideas or a decline in the impact of these ideas on growth – we place firms center stage and study each relationship separately. We first focus on evidence that speaks to researchers' ability to generate new ideas. For our purposes as micro-economists, it is useful to define a simple measure of

[8] Romer (1990) notes that "[l]inearity in A is what makes unbounded growth possible," and that researchers would shift towards manufacturing as A grows, thereby lowering growth, if equation (1) were concave in A. Note that in Jones (1995), the $1-\beta$ above is instead denoted by ϕ. Here we follow the notation in Bloom et al. (2020).

[9] Specifically, they measure $\beta = \frac{g_{H_A}}{g_Y}$, where g denotes growth rates.

[10] Jones (2009) provides microeconomic evidence that researchers' productivity, measured as their age at first invention and size of teams both rise as knowledge grows. He links these patterns to aggregate growth via a model in which research productivity declines, but he does not provide direct evidence on how such changes map to productivity growth.

research productivity that is readily observable in the data and free of any model or functional-form assumptions. In particular, we propose

$$\text{Research Productivity} \coloneqq \dot{A}_t/H_{At} = \frac{\text{new patents}}{\text{research expenditure,}}$$

which we will measure using patents per real research expenditures. Declines in such research productivity motivated a number of early papers in the semi-endogenous growth literature, including Jones (1995) and Kortum (1997). This definition also shows how research productivity in Romer (1990) is increasing in both the productivity parameter δ and the aggregate stock of ideas ($\dot{A}_t/H_{At} = \delta A$), whereas in Jones (1995) there are diminishing returns to aggregate ideas when $0 < \beta < 1$ ($\dot{A}_t/H_{At} = \delta A^{1-\beta}$).

As a second metric on how researchers' ability to generate new ideas has evolved over time, we estimate the elasticity of patents to research inputs. To do so, we adapt equation (5) to:

$$Patents_{ft} = \delta ResearchInputs_{ft}^{\eta_t} A_t^{1-\beta}, \tag{6}$$

where η_t is the elasticity of firm f's patents (ideas) to its research inputs (H_A) up through year t. We measure research inputs using traditional R&D expenditures, as well as with more comprehensive measures based on firm payroll in innovation-related establishments. Although the exponent on research inputs is absent in Romer, both Jones (1995) and Bloom et al. (2020) consider its impact, referring to it as the 'stepping on toes' effect. The distinction here is that we estimate a firm-level elasticity, whereas they consider diminishing returns to aggregate research efforts. An estimate of this patent elasticity, η_t, less than one implies diminishing returns to R&D efforts for a given firm in a particular time period that would in turn dampen the impact of additional researchers on aggregate growth. In particular, even if $\beta = 0$, the steady-state growth rate of ideas would be lower than the growth rate of research inputs (H_A), with diminishing returns to researchers creating a gap between the growth of research inputs versus output growth rate. Crucially, however, and as noted in Romer (1990) and Jones (1995), unbounded growth still occurs with a fixed population when $0 < \eta_t < 1$ as long as $\beta \leq 0$.

In our formulation, the term $A_t^{1-\beta}$ is the aggregate stock of non-depreciated ideas (patents) that affects all firms equally. Treating A_t as an aggregate rather than firm-specific stock is crucial for capturing the main mechanism in Romer: ideas are (at least partially) non-rival and non-excludable, such that any given firm's efforts to generate new ideas will depend on the total stock of ideas in the economy. As such, it will be captured by year fixed effects in our empirical implementation. The year fixed effects also capture any changes in spillovers across firms (Bloom et al., 2013), effects of declining public R&D expenditures on innovation (Bergeaud et al., forth), and the discovery of new technology "classes," e.g, the PC-internet revolution of the late 1980s and 1990s (Ribeiro, 2025). It is worth noting that if Romer's productivity parameter (δ) were time-varying, it would similarly be captured by the year fixed effects. We will therefore estimate and plot these year fixed effects as a third way to assess whether ideas may be getting harder to find, *independent* of firm's research inputs. Although we might have included the firm's own past realizations of ideas (A_{ft}) as a separate term,

in practice we will subsume various potential lag structures for the firm's research-input flows using a depreciated stock of its research inputs, which is empirically (and theoretically) correlated with its patent stock, thus making it infeasible to identify separately from its research inputs.

We next consider the second fundamental relationship in Romer: how ideas map to growth. If taken literally, Romer predicts growth solely through firm entry, with each new idea generating a new input producer. The fundamental mechanism at play, however, is that a new idea increases output from a fixed amount of labor and human capital. To capture this spirit of Romer, we study how a firm's growth rate relates to changes in its stock of ideas by adapting equation (3) to

$$Growth(Y_{ft}) = \mu_t Growth\left(PatentStock_{ft}\right) + \alpha Growth(L_{ft}) + (1-\alpha)Growth(K) + \phi A_t, \quad (7)$$

where Y_{ft} is real sales, $PatentStock_{ft}$ is the firm total patents up to year t, L_{ft} is firm employment in t, and K_{ft} is firm capital in t. In this adaptation, we combine labor services (L) and human capital (H_A) into L. We control for changing labor at the firm level in order to isolate the relationship of patents on growth. Because we lack observable measures of capital for non-public firms outside manufacturing, we control for changes in firm-level capital using firm size and age bins. An estimate of $0 < \mu_t < 1$ indicates positive, though diminishing, returns to ideas for a particular firm and time period. We also posit the potential for the aggregate stock of ideas to influence an individual firm's growth by including A_t. Note that, as with idea creation in equation (6), any changes in the relationship between firm growth and the aggregate stock of ideas, or in the effect of knowledge spillovers on firm growth, are captured by year fixed effects.

Our firm-level approach allows us to pinpoint the mapping from idea inputs (R&D) into idea outputs (patents), and to assess whether and how the latter maps to firm growth. In our current specifications, all firms are affected equally by the growing aggregate stock of ideas (A_t), so any knowledge spillovers are captured by year fixed effects. While we might have skipped the intermediate 'idea' step by simply estimating the relationship between R&D and growth, our approach allows us to identify where the break occurs. In addition, R&D entails substantial risk and uncertainty. By studying the relationship between idea and output growth directly, we mitigate any noise due to the uncertainty in translating R&D inputs into ideas.

3 Data

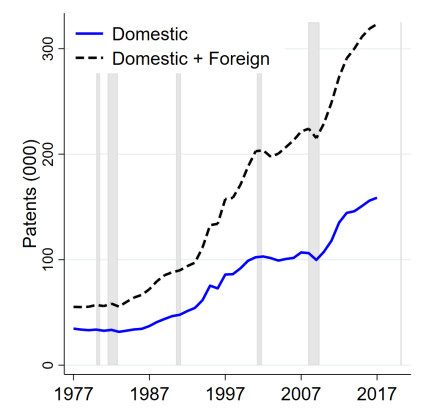
In this section, we explain the construction of our new 45-year firm-level panel dataset of US firms' patents and research-input use.

3.1 Patent Data

We measure a firm's output of ideas using patents. The US Patent and Trademark Office's (USPTO) PatentsView (PV) database provides information on the identity and location of all granted US patents from 1977 through 2021 and their corresponding assignees and inventors, dates of application

and granting, and citations. Figure 2 reports the relative prevalence of foreign and domestic granted patents by application year. Since the median patent has a five-year lag between its application and grant date, we use patents through 2016 to mitigate truncation at the end of the period.[11] We restrict our analyses to "domestic" patents, which we define as patents with at least one assignee located in the United States. We exclude "foreign" patents because the LBD contains no information on firms without US establishments, and we would potentially overstate researcher research productivity by including foreign inventions in their idea counts.[12] All subsequent references to patents refer to domestic patents, unless otherwise specified.

Figure 2: Domestic vs Foreign Patents by Application Year



Source: PV and authors' calculations.
Notes: Figure reports total (black dashed line) and domestic (blue solid line) patents in the PV dataset over our sample period. Domestic patents are defined as having at least one assignee with a US location.

Griliches (1998) highlights patents as a measure of innovation, but notes two limitations relevant here. First, not all inventions are patented, for instance because firms protect them as trade secrets instead. Since this paper relies on patents as a measure of innovative output, our analysis may underestimate the extent to which research inputs generate ideas. Second, patents differ greatly in their quality or economic impact (Bessen, 2008). To address potential variation in patent quality and impact, we rely on two commonly used measures: forward "external" citations and breakthrough patents.

To compute external citations, we follow the recommendation from Jaffe and de Rassenfosse (2017) and sum the total number of forward citations received by a patent within the first five years of the patent's grant date.[13] We exclude the firm's own "self" citations, since these may be more related to

[11] The mean lag between patent application and granting is 2.6 years, with median, 90th, 95th, and 99th percentiles at 2, 5, 6, and 8 years, respectively. Thus, about 90 percent of patents applied for in 2016 have been granted by 2022.

[12] Our domestic patents thus include patents by foreign firms operating in the United States as long as they have a US establishment listed as an assignee. Our goal is to exclude patents without US inventors or R&D inputs, since those likely reflect knowledge foreign multinationals accumulate in their home countries, which we do not observe. Although US multinationals may benefit from research activities abroad, which we also do not observe, Bilir and Morales (2020) show that US MNEs' foreign R&D has negligible effects on their domestic outcomes.

[13] Windowing ensures that all patents have the same time to accumulate citations. We classify citations from patents

patent thickets aimed at preventing competition than to increases in new ideas (Cohen et al., 2000). We measure breakthrough patents from Kelly et al. (2021), hereafter KPST. They estimate patent "novelty" as the ratio of a patent's forward textual similarity to its backward textual dissimilarity with other patents over five-year intervals. In our analysis, we define breakthrough patents as those whose novelty is in the top five percent across all patents from 1836 to 2021, net of year fixed effects.[14] Since citations and breakthroughs are both forward-looking, these measures are available through 2011.

We also consider a fourth patent measure based on the stock market's expected future value of a newly granted patent. Kogan et al. (2017) – hereafter KPSS – estimate patent "value" as the average abnormal returns in a firm's stock market price on the day it receives the patent grant.[15] These estimates are available through 2016 for the subset of publicly traded firms. Patent value is distinct from our other measures because it combines the two relationships we seek to estimate in equations (6) and (7). Patent value also reflects the expected *private* returns to an idea, rather than its societal value. For instance, a patent for an important scientific breakthrough that garners numerous citations may not have a high estimated value unless its benefits can be captured by the firm. Alternatively, a patent with little scientific merit may nevertheless be of high value if, for example, it can be used to restrict the firm's competition.

3.2 US Census Establishment- and Firm-level Data

We construct a firm-level panel of employment, payroll, sales, and R&D expenditure by combining multiple datasets from the US Census Bureau. We start with the Longitudinal Business Database (LBD), which provides employment, payroll, industry, and geography for all private, non-farm employer establishments from 1976 to 2021 (Jarmin and Miranda, 2002; Chow et al., 2021). We supplement the LBD with establishment-level sales using the quinquennial Economic Census data conducted in years that end in 2 or 7.[16] We restrict the analysis to establishments that are in scope for the Business Dynamics Statistics (BDS) data, which ensures that we use research-quality observations and have sample that aligns with published official statistics on the US economy. This limitation excludes establishments in Government and those that the Census Business Register considers out of scope. We also drop Public Administration (NAICS 92) and Agriculture (NAICS 11) establishments, as well as establishments with zero payroll (i.e., "non-payroll active" establishments).

We aggregate the establishment-level data to the firm level using the Census Bureau's cross-sectional firm identifier (*firmid*), which we correct for spurious longitudinal breaks that occur when firms transition between single and multi-unit status. A benefit of the Census establishment-level data is that it records every establishment's primary industry code in each year. We construct annual measures of each firm's employment and payroll across sectors using the vintage-consistent North

for which we do not have a matched firm identifier (about 8 percent of patents) to be an external citation.

[14] Their updated dataset includes over 11.4 million patents granted from 1836 to 2022. They define a "breakthrough" patent as one whose novelty is in the top 5 or 10 percent across all patents over that interval, net of year fixed effects. Intuitively, these patents represent "breaks" with the past that are also widely emulated going forward.

[15] Values are expressed in real terms using 1982 dollars. The most recent vintage of these data covers over 3.1 million patents granted between 1926 and 2022 that can be matched to publicly traded firms.

[16] The Economic Censuses include the Census of Manufactures, Wholesale Trade, Retail Trade, Services, Finance, Insurance, and Real Estate, Construction, Transportation, Communication and Utilities, and Mining.

American Industrial Classification (NAICS) codes developed by Fort and Klimek (2018). These codes ensure that we assign establishments consistently to sectors and industries over the period, even as the underlying classification systems change. We assign firms to their primary industry in each year based on the majority of their payroll. The mix of firm employment across establishments in different industries is also crucial for the construction of alternate measures of innovation inputs, described below. We follow the BDS and define firm age as the difference between the current year and the first year of positive employment of the firm's oldest establishment in the year that the $firmid$ first appears.

The majority of past studies on research efficiency measure innovation inputs using R&D expenditures. We therefore compile R&D expenditures over the 1977 to 2021 period for the subset of firms in the National Science Foundation's R&D surveys, collectively referred to as RADS. These surveys provide R&D expenditures for a rotating sample of approximately 45 thousand firms each year, though the sample size and frame varies over time. We aggregate R&D expenditures to the firm level and merge them to the LBD panel (see Appendix D for details).

3.3 Compustat Firm Panel

Since a large portion of past work on firm innovation and researcher efficiency relies on publicly traded firms (a subset of the firms in the LBD), we construct a similar firm-level panel using Compustat. Compustat is produced by S&P Global Market Intelligence and contains financial information for publicly traded firms derived from companies' Security and Exchange Commission filings, including 10-K's and 10-Q's. Compustat contains a wide range of information about firms' global attributes and performance, including their global sales, employment, cost of goods sold (COGS), Selling, General and Administrative (SG&A), and R&D expenses.

3.4 New Measures of Research Inputs

A key ingredient necessary to estimate research efficiency is a measure of knowledge or innovation inputs. While R&D expenditure is a natural starting point, the rotating nature of the R&D surveys, changes in their sample-selection criteria, and their tilt towards large manufacturing firms pose serious limitations on the sample of firms we can study. Past work also documents "missing R&D" expenditures among public firms that can bias estimates of the impact of R&D on patenting (Koh and Reeb, 2015).

To mitigate potential biases from sample selection or variation in R&D reporting over time, we propose three new measures of 'research inputs' in the LBD panel based on a firm's payroll in innovation-related establishments. Our most narrow measure calculates a firm's wage bill in R&D labs (NAICS 5417). While such labs are appealing due to their clear focus on innovation, they may miss considerable research efforts to the extent that plants whose major activity is not R&D nevertheless contain workers engaged in these activities. We therefore construct a second measure based on the firm's wage bill in Professional, Scientific, and Technical Services as well as Management establishments (NAICS 54-55). These sectors cover establishments engaged in high-skill, more

knowledge-intensive tasks such as engineering and computer system design.[17] Finally, we provide an upper bound on firms' R&D employment expenditures using their total payroll across all establishments. Although clearly a crude measure, total payroll covers all potential patenting firms in our sample.

For Compustat firms, we use both capitalized selling, general and administrative expenses (SG&A) – a broad measure of innovation inputs that is often used in the construction of firms' intangible assets – as well as their reported R&D expenditures.

We deflate the R&D variables, as well as all other nominal variables, e.g., firm sales and payroll in the LBD, using the CPI. While some researchers deflate these series using the average skilled-worker wage (e.g., Bloom et al., 2020), we view the CPI as conservative for our purposes insofar as it captures increases in worker skill and ability that would be deflated away using skilled-worker wages.

3.5 Matching Patents to Firm-level Panel Data

We merge the USPTO patent-assignees to firm names and geography in the Census Bureau's Business Register (CBPBR). The CBPBR is built from the universe of administrative Internal Revenue Services payroll tax filings and provides the frame for the LBD. The resulting crosswalk between the USPTO and CBPBR, referred to as BDSPF-Long and described in detail in Appendix A, makes four contributions relative to past matches: it spans the longest time period of existing patent-to-$firmid$ matches (1977 to 2021), it prioritizes matches' longitudinal consistency and accuracy thereby producing high match and precision rates (92 and 94 percent), it is available along with its underlying code to researchers in the FSRDC network with approved projects, and it will be updated annually.[18] To the best of our knowledge, our bridge is the first to incorporate longitudinal consistency in $firmid$ matches when combining an external dataset to the Census data. It is also used to construct annually updated public-use tabulations of patenting firms' business dynamics statistics.[19]

Although the BDSPF-Long has higher match rates than previous efforts, it is important to note that the match rates of all bridges between patents and firms decline over time, in our case from 95 percent in the early 1970s to 90 percent in the late 2010s (Appendix Figure A.1). This decline is driven by increases in the shares of patent-assignee records that do not match to the CBPBR at all, as well as those that match based upon fewer and noisier characteristics, trends that suggest patents are increasingly granted to small (i.e., single-establishment) firms.[20] Since we use patents as our measure

[17] NAICS 54 encompasses nine 4-digit NAICS industries: legal (5411), accounting (5412), engineering (5413), design (5414), computer system design (5415), consulting (5416), R&D (5417), advertising (5418), and other (5419). We include "Management" establishments (NAICS 55), since they often employ R&D workers, and because establishments that perform two or more Professional, Scientific, or Technical Services for other establishments of their firm are classified in this sector.

[18] The crosswalk from Balasubramanian and Sivadasan (2011) spans 1975 to 1997, Kerr and Fu (2008) spans 1976 to 2000, and the BDS-PF Triangulation bridge by Graham et al. (2018) begins in 2000. Appendix A provides a comparison of BDSPF-Long to these other matches. We match 92 percent of domestic patent-assignees to CBPBR. Using the BDS-PF Triangulation bridge A1 high-quality matches as "truth", we find that the ratio of true positives to the sum of true positives and false positives (i.e., precision), is 94 percent.

[19] The BDSPF tabulations are available at `https://www.census.gov/programs-surveys/bds.html`.

[20] See Appendix Figure A.2. As illustrated in Appendix Figure A.3, match rates would have remained at about 94 percent if the raw match-pass rates had held constant. Finding a match between a patent assignee and a single-establishment firm is probabilistically more challenging because such firms have only one observation in each year in

of ideas, the secular decline in match rates is most likely to exert a downward bias in our estimates of researcher efficiency.

We use the BDSPF-Long (patent to CBPBR) crosswalk to assign patents to *firmid*s in our LBD panel. We create a separate match of US patents to the publicly traded firms in Compustat by combining two recent patent-to-firm mappings developed by Kogan et al. (2017) and Dyevre and Seager (2024). We exploit these two mappings to increase the overall number of patents we can match to Compustat, and to cross-validate patents encompassed by each approach. See Appendix B for further details.

In both panels, we match patents to firms by grant year, but use the patent *application* year in our analyses so that they are most proximate in time to innovation investments.[21] As noted at the beginning of this section, because of the typical gap between patents' application and grant years, our analysis sample period runs from 1977 to 2016 for patent grants, and 1977 to 2011 for external citations and breakthrough patents.

We report match rates for patent grants and breakthroughs by semi-decade in Appendix Table A.8.[22] As the firms in our LBD panel are a subset of the entities in the CPBPR (see Section 3.2), the match rate for the LBD is lower than quoted above – 89 percent of all domestic patents in the first semi-decade versus 85 percent by the final semi-decade. We note, however, that the beginning-to-end 4 percentage point decline in the match rate in the LBD is slightly less than the 5 percentage point decline in the overall crosswalk rates in the CBPBR. By contrast, the match rates in Compustat start lower and decline relatively more over the same period, falling from 63 to 55 percent. The LBD panel thus covers 26 to 34 percent more patents than Compustat, with steadier match rates throughout.

Match rates for breakthrough patents are also higher in the LBD panel relative to Compustat. The LBD panel covers 91 percent of breakthroughs in the first period and 87 percent by the end, whereas Compustat coverage is about 73 to 72 percent in those periods. Compustat coverage is also more volatile over the four decades: in the 1987 to 1991 period, its coverage falls to just 60 percent of breakthrough patents, while the LBD panel is steady at 87 percent.

The share of patent activity in Compustat may decline for at least two reasons. First, as a higher share of firms chooses to stay private, Compustat's coverage of economic activity will naturally decline (Doidge et al., 2017; Schlingemann and Stulz, 2022). Second, as we will show in the next section, the composition of US patenting firms shifts towards smaller, non-manufacturing firms, both of which are less likely to be in Compustat. Regardless of the underlying reasons for the decline, the lower and more volatile match rates in Compustat suggest that those data may not provide as accurate a representation of US research efficiency as the LBD, since falling coverage may manifest as falling efficiency.

the CBPBR. Multi-establishment firms, by contrast, have a separate observation for each establishment. In addition, the Census collects more detailed and frequent information for multi-unit firms.

[21] Our LBD panel will also be missing any matches in the crosswalk for which the matched LBD firm is not active in the patent's application year, though we allow for pre-birth patenting in the growth regressions in Section 6 when constructing firms' patent stocks.

[22] We do not report rates for external citations, since these are defined only for matched firms, or for patent value since this measure only exists for publicly traded firms.

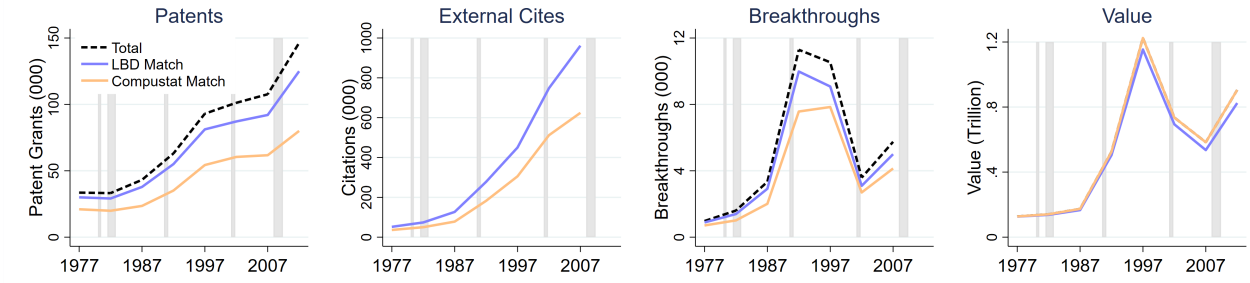
4 New Facts on US Patenting Activity

In this section, we exploit our long time series on the universe of private-sector firms to provide 4 new facts about US patenting activity that are relevant to understanding the evolution of US researcher output, efforts, and efficiency.

4.1 Trends in US Patent Activity

Figure 3 depicts domestic patent grants, external citations, breakthroughs, and value in the LBD and Compustat panels. To minimize disclosure burden, each point contains a five-year forward-looking average, e.g., 1977 is the average from 1977 to 1981. In the first panel, the line for "Total" is total *domestic* patents grants in PatentsView, *i.e.*, the trend for domestic patents displayed in Figure 2. There are two key messages from this figure. First, patent grants and external citations rise over the period, and more so among the universe of private-sector firms relative to the subset of publicly traded firms. Second, breakthrough patents exhibit a dramatic surge in the early 1990s, with a sharp decline after 2002; they begin to recover in the last 2007 to 2012 period. Patent value exhibits a similar inverted-u shape, though its peak occurs about five years later, right before the dot-com burst. Since value is based on stock market returns, it is only available for publicly traded firms. It also mixes the two relationships we seek to distinguish: that between R&D inputs and ideas versus ideas and growth. We therefore focus on the first three measures for the remainder of the paper.

Figure 3: Patent Activity in the LBD and Compustat Panels



Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, Compustat and authors' calculations.
Notes: Figure compares LBD- and Compustat-matched domestic patent grants, external citations, breakthrough patents, and patent value to their totals in the PV data. Patents are defined as domestic if at least one assignee has a US location. Each point represents an average across five-year semi-decades, e.g., the data point for 2012 is the average across 2012 to 2016. Shading represents US recessions. The last point for breakthrough and external citations is 2007 (the average of 2007 to 2011) as these measures are forward-looking. Legend for all panels is in the first panel. External citations are defined only for matched patents and value is defined only for public firms, so no totals are reported. All series are plotted by patent application year.

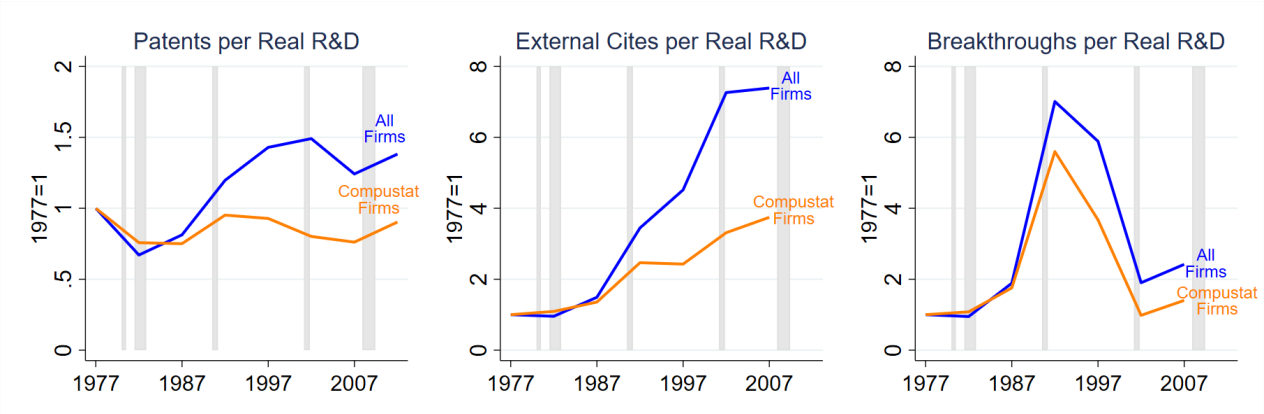
4.2 Patents per R&D Input

A natural first step to analyzing whether ideas are getting harder to find is to measure research productivity, *i.e.,* average patents per R&D input. Indeed, some of the key growth papers in the 1990s developed models to match a salient fact on declining patent productivity: the number of patents per R&D researcher (or R&D dollar) declined both in the US and across countries starting

in in the 1970s (Evenson, 1984; Kortum, 1993; Jones, 1995; Kortum, 1997). We replicate past work to depict how such research productivity evolves in the LBD and Compustat firm panels and over a longer time period.

Figure 4 presents research productivity – patent activity per real R&D expenditure flows over time in the LBD and Compustat. Prior work divides R&D expenditures by R&D worker wages to infer the number of researchers in these calculations. We deflate expenditures by the CPI to allow for heterogeneity in researcher ability that would likely be captured in greater real expenditures. We again minimize Census disclosure burden by reporting results by semi-decade

Figure 4: Research Productivity in the LBD and Compustat



Source: PV, LBD, KPST, RADS, BDSPF-Long, DS, Compustat and authors' calculations. *Notes:* Figure reports patents, breakthroughs, and external citations per real R&D expenditure for the firms tracked by the LBD and Compustat panels. Ratios are averages across 5-year semi-decades from 1977 to 2012, and these averages are indexed to 1 in the first, 1977 to 1981 semi-decade. Patents are assigned to firm by application year. In the LBD panel, the set of firms included in the ratio is restricted to those in the R&D surveys in each year.

The first panel in Figure 4 depicts total granted patents per real R&D dollar. Among the universe of firms in our LBD panel, this measure falls from the 1977 to the 1981 semi-decade (i.e., the average of 1977 to 1981 is higher than the average from 1982 to 1986), grows steadily until the 2002 semi-decade, and then falls and almost fully recovers by the 2012 to 2016 period. Most notably, and in contrast to past evidence, research productivity rises over time: by the last period, it is almost 1.5 times what it was in the late 1970s. The first panel of Figure 4 also depicts patents per R&D dollar among Compustat firms. While publicly traded firms' research productivity tracks overall efficiency quite closely for the first decade, there is a sharp divergence in the 1990s. In sharp contrast to the same measure among the universe of private firms in the LBD, research productivity for Compustat firms is lower in the final semi-decade than it was in the late 1970s.

The second panel of Figure 4 shows that external citations within 5 years of the grant date per R&D input grow fairly steadily until the 2007 semi-decade, at which point they flatten out. Breakthroughs per R&D dollar have the same basic inverted-U shape as their overall trend, with a massive surge in efficiency during the 1990s among both LBD and Compustat firms. Despite this surge and subsequent decline, breakthrough efficiency among both sets of firms finishes the period higher than it started. There is no evidence of a secular decline in these patent activities per R&D dollar among LBD or Compustat firms over our period. We summarize these results in our first fact:

Fact 1. *Research productivity – measured as patent grants, external citations, or breakthroughs per real R&D dollar – all rise over the last four decades for the universe of firms. By contrast, Compustat firms' patent grants per R&D dollar are lower at the end of the sample period than the beginning. Their citations and breakthroughs per R&D dollar, however, both rise, albeit less sharply than for the universe of firms.*

Fact 1 demonstrates the importance of examining the universe of firms for a long time period using multiple measures of idea output when drawing conclusions about long-run trends in researcher efficiency. A number of papers develop models with diminishing returns to research efforts to explain the pattern of declining patents per R&D dollar over multiple decades evident in Compustat. However, the same research productivity metric is generally rising among the universe of private-sector firms and is flat or rising for other measures of patenting among Compustat firms.

4.3 Changing Composition of Patenting Firms

To understand the evolution of US patenting and the sharp divergence in research productivity documented above, we decompose LBD firms' patenting activity along three dimensions: the firm's major sector of activity (based on payroll), the firm's size according to employment, and the firm's age. We show how these activities have changed over time with separate decompositions for each decade from the 1970s to the 2010s, where the first and last decades consist of 1977 to 1979 and 2010 through 2016, respectively.

The top row of Figure 5 reveals a steady decline in manufacturing firms' shares of patent grants, citations, and breakthroughs, which decrease from above 70 percent in the 1970s to 30, 23, and 12 percent, respectively, by the 2010s (see Appendix Table A.12 for numbers underlying the figure). By contrast, firms in the Information (NAICS 51), Professional Services and Management (NAICS 54-55) and Wholesale (NAICS 42) sectors exhibit considerable growth in patenting shares over time. Firms in Professional Services and Management account for about 30 percent of each patent activity by 2010s. Firms in Information, which includes software, telecommunications, and data processing, are the most prominent breakthrough patenters in the 2000s and 2010s, accounting for 37 percent of breakthroughs in the last decade. We summarize these stark patterns in a second fact:

Fact 2. *US patenting was dominated by manufacturing firms from 1977 until 2000. By the 2010s, manufacturing firms' account for only 30 percent of patent grants, 23 percent of external citations, and 12 percent of breakthrough patents. By contrast, firms in Information, Professional Services, Management, and Wholesale grow their patent activity shares over the 1977 to 2016 period.*

Fact 2 demonstrates that a complete evaluation of US researcher efficiency must expand beyond the manufacturing sector. The rise in patenting among Wholesale and Professional Services firms resonates with work on the increasing prevalence of factoryless goods producers – firms that design goods and coordinate the production process, but outsource physical transformation tasks to other firms that are often in other countries (Bernard and Fort, 2015; Kamal, 2023; Fort, 2023). Firms with

Figure 5: Patent Activity by Firm Sector, Size, and Age (LBD Panel)



Source: PV, LBD, KPSS, KPST, BDSPF-Long, and authors' calculations.
Notes: Figure provides a breakdown of patent grants, external citations, and breakthroughs by LBD firms' major 2-digit NAICS sector, size, and age. Firms in the LBD cannot be observed prior to 1977. As a result, the majority of firm observations have left-censored age in the early years of the panel.

Management as their primary sector may reflect US multinationals' strong role in US innovation, given their disproportionate employment shares in that sector (Kamal et al., 2022). The evidence here indicates that these firms, some of which may have transitioned from the manufacturing sector (Ding et al., 2022), are increasingly important for US innovation, but likely to be missed by studies with the traditional focus on manufacturing.

The middle row in Figure 5 depicts patent activity shares by firm size categories. The first salient

message is the declining dominance of mega-firms – those with more than 10 thousand US employees – in US patent grants and citations. In the late 1970s, these firms accounted for 58 percent of both measures. By the 2010s, their share of grants and citations had fallen over 15 percent, to 48 and 49 percent, respectively. Although all other firm-size categories grow their grant and citation shares over the period, small firms (1 - 19 workers) have the most notable growth. Their shares of patent grants and citations rise 3 percentage points, or about 40 percent, over the period. This shift in patenting from large to small firms is surprising since aggregate US employment shifted away from small firms (4 percentage point decline) towards mega-firms (3 percentage point increase) over the same period (see Appendix Table A.14).

Mega-firms' share of breakthroughs evolves differently, as it is U-shaped over time: larger in the 1970s and 2000s, but smaller in the 1980s and 1990s.[23] Instead, firms with 100 to 500 employees almost double their share of breakthroughs in the 1980s, while in the 1990s firms with 2,500 to 10 thousand workers jump from 10 to 16 percent of these breakthroughs. By the 2010s, mega-firms have regained their dominant share, accounting for 67 percent of breakthrough patents. These changes likely reflect a combination of strong growth by a select set of firms with such patents as they transition across size bins, along with potential changes in the types of firms that innovate successfully in different periods. For example, the 1990s witnessed an explosion of electrical engineering patents during the personal computer and internet boom (see Appendix Figure A.10), with a small subset of these firms growing into the tech giants that persist today.

The bottom row of Figure 5 illustrates how patents are distributed across firms of different ages. Since we construct firm-age bins using the minimum year of the firm's oldest establishment, age is left-censored in different years for different categories making time-series comparisons difficult. We therefore focus on comparing shares within an age group across patent activities.

There are several notable patterns about patenting and firm age. First, younger firms (0-10 years old) have a disproportionate share of citations relative to their granted patents in each of the decades for which we observe their age. In the 1990s, 2000s, and 2010s, they account for 15, 14, and 13 percent of grants versus 19, 18, and 17 percent of cites. By contrast, they have low shares of breakthroughs relative to total grants. Second, 11-15 year-old firms have disproportionate shares of breakthroughs throughout, but particularly in the 2010s. Indeed, their share of breakthroughs rises from 8 to 14 percent from the 2000s to the 2010s (there is no age truncation for this category in these years). The data thus reveal that firms born around 1995 to the early 2000s, just after the breakthrough boom and before the internet bubble burst, are particularly strong innovators about 10 to 15 years later. We summarize these findings on firm size and age in the following third fact:

Fact 3. *Mega-firms (10k+ employees) dominate US patenting until the 1990s, but their shares of patent grants and citations fall to less than 50 percent by the 2010s. By contrast mega-firms account for 70 percent of breakthrough patents in the late 1970s and 67 percent in the 2010s, though only 59 percent in the 1990s. This U-shaped pattern of breakthrough patent shares by mega-firms is concurrent with a rise in breakthrough shares by 11-15 year-old firms from the 2000s to the 2010.*

[23] Braguinsky et al. (2025) define novel patents as those that combine two technology classes for the first time and similarly find that mega-firms' importance in such patents exhibits a U-shaped pattern over this period.

These changing patterns of firm size and age in US patenting highlight the need to control for both factors when evaluating research output and its impact on growth. They also suggest that innovation may come in bursts from new discoveries, followed by growth and consolidation by a subset of those firms with breakthroughs. The recovery of breakthrough shares by a small set of mega-firms may also signal a changing market structure for new technology fields in which innovators themselves grow, even as they acquire other firms and stifle competition and exclude rivals from growing using their ideas.

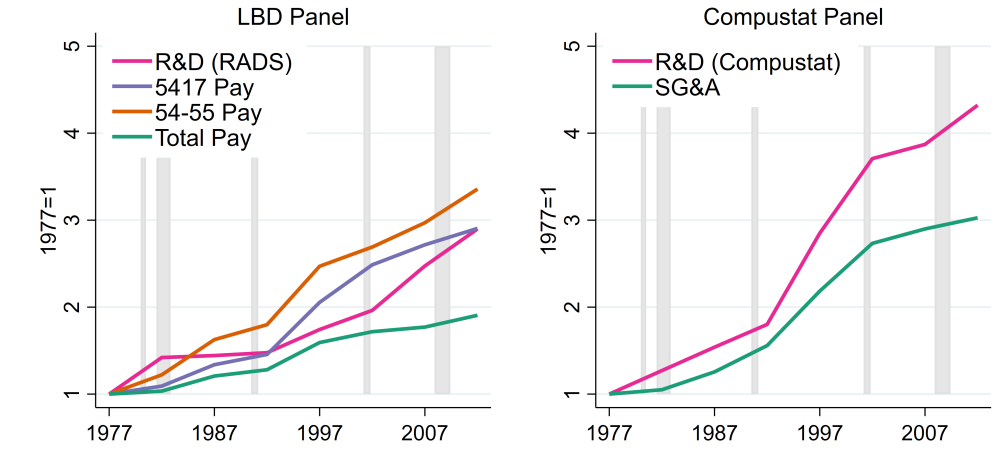
4.4 Trends in Patent Coverage by firms with Research Inputs

In the next section, we will estimate the elasticity of patents to R&D inputs. Given the changing composition of US patenters towards smaller, non-manufacturing firms, we want to ensure as complete and accurate measures of these R&D inputs over time. On the one hand, the R&D survey frame tends to target larger, manufacturing firms. On the other hand, the frame for the R&D surveys has expanded over time to include more firm types. Either of these changes may lead to biased measures of how R&D expenditure has changed over time. To minimize the potential bias from mismeasured R&D inputs, we analyze the evolution of the various research-input flows we propose in Section 3.4 and show how the fraction of patents covered by firms with these inputs evolves over the period.

Figure 6 depicts average aggregate innovation-input flows across firms in the LBD (left panel) and Compustat (right panel). As above, each point contains a five-year forward-looking average, *e.g.*, 1977 is the average from 1977 to 1981, and each series is normalized to to 1 in the first period. As indicated in the figure, R&D expenditure among LBD firms rises substantially, by a factor of 2.9 over the entire period. Reassuringly, R&D lab (NAICS 5417) payroll growth aligns closely with R&D expenditure, though it grows more steadily throughout. Payroll in Professional Services and Management establishments grows the most, rising by a factor greater than 3. By contrast, total real payroll does not quite double over the 40-year period. The right panel shows that Compustat firms' R&D expenditure grows more quickly (by a factor of more than 4), while SG&A expenditure tracks fairly closely with the LBD payroll in Professional and Management Services. The lower patents per R&D dollar in Compustat versus the LBD summarized in Fact 1 is thus due to both the numerator growing more slowly (see Figure 3) and the denominator growing more rapidly among the subset of publicly traded firms.

We also analyze how well firms with each research-input flow cover total US patent grants. Figure 7 shows that firms with R&D expenditures in the RADS surveys account for about 65 percent of our matched LBD patents in the 1970s and 1980s decades, but just 59 percent in the 1990s through the 2010s. This decline coincides with the compositional shifts away from manufacturing and larger firms documented in Facts 2 and 3, and highlights the potential limitations of the R&D surveys in measuring research inputs. Firms with R&D labs also exhibit decreasing coverage. They account for the majority (54 percent) of patents in the 1970s, but only 48 percent by the 2010s. By contrast, Figure 7 shows that firms with Professional Services and Management establishments cover 77 of patents in both the 1970s and the 2010s, thus raising our coverage and improving its stability in the

Figure 6: Research-Input Flows



Source: PV, LBD, BDSPF-Long, DS, RADS, Compustat and authors' calculations.
Notes: Figure reports average annual real aggregate research-input flows across firms in the LBD (left panel) and Compustat (right panel) by semi-decade. In the LBD panel, research inputs are firms' RADS R&D expenditure, R&D lab (NAICS 5417) payroll, Professional Services and Management (NAICS 54-55) payroll, and total payroll. Series are normalized to 1 in the first semi-decade. Reporting is done by semi-decade to minimize disclosure concerns. A data point for 2012, for example, is the average flow across firms from 2012 to 2016.

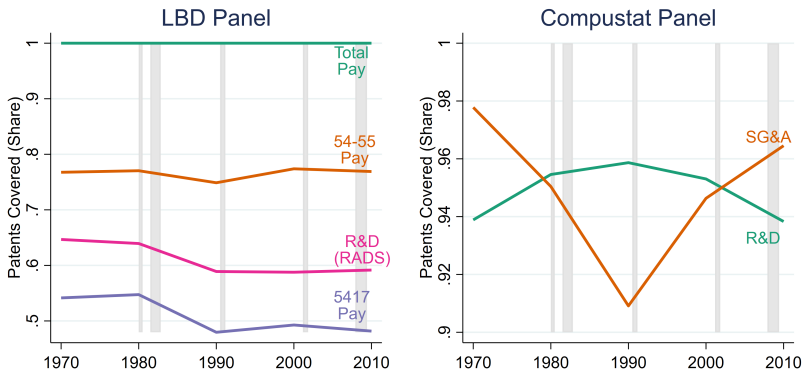
time series.[24] We summarize these findings in the last fact:

Fact 4. *The share of patents covered by firms in the Census R&D survey falls from 65 percent in the late 1970s to 59 percent in the 2010 decade, with the most substantial decline occurring in the 1990s. The share of patents covered by firms with employment in Professional, Scientific, and Technical Services and Management establishments hovers at about 77 percent throughout.*

To recap, this section documents a rise in patents per real R&D expenditure among the universe of firms, even as such research productivity has declined among the subset of publicly traded firms in Compustat. This divergence in efficiency coincides with a changing composition of patenting firms in the United States. A rising share of patents is now done by smaller non-manufacturing firms, which aligns with decreasing coverage of US patenting activity in Compustat. The share of patents covered by firms in R&D surveys also declines over time, prompting our development of multiple new measures of firms' research inputs. We use those measures to estimate the elasticity of patents to research inputs in the next section.

[24] Fifty-four percent of the firm-year observations in our Compustat extract report missing R&D. This share drops to 19 percent among firms with patents. These missing observations exert little influence on the trends reported in this paper because they tend to occur among firms with few patents. For further discussion of missing R&D and other research input observations, see Appendix Section C and Appendix Table A.9.

Figure 7: Patent Coverage by Firm Research-Input Flows



Source: PV, LBD, BDSPF-Long, DS, RADS, and authors' calculations.
Notes: Figure reports the share of patents in our LBD and Compustat panels covered by firms with noted research-input flows. For the LBD panel, these flows are R&D from RADS, payroll in R&D labs (NAICS 5417), payroll in Professional, Scientific, and Technical Services or Management establishments (NAICS 54-55), or overall payroll. Coverage for the latter is 100 percent in all years as all LBD firms with patents have payroll. For Compustat, the knowledge flows are SG&A and R&D.

5 Patent Elasticity Estimates

In this section, we exploit our firm panel to estimate the elasticity of patents to R&D inputs, controlling for both firm and time fixed effects. The patent elasticities provide a second measure of how research efficiency evolves over time. The time fixed effects provide a third metric that captures changes in average firm patenting that is independent of variation in their R&D expenditure.

5.1 Empirical Specification

We estimate time-varying elasticities of a firm's patents to its research inputs by adapting equation (6) to

$$Y_{ft} = exp\left(\sum_{j=1977(5)k} \eta_j d[j]_{ft} \times ln(ResearchInputs_{ft}) + \sum_{j=1977(5)k} \gamma_j d[j]_{ft} + \gamma_f\right) + \epsilon_{ft}, \quad (8)$$

where Y_{ft} is firm f's patent output – patent grants, breakthroughs, or external citations – in year t and $ResearchInputs_{ft}$ captures the firm's R&D inputs (i.e., H_A in Romer's notation). We interact the firm's research inputs with dummies $d[j]_{ft}$ that are equal to 1 for years t in semi-decade j. For example, $j = 1977$ encompasses years 1977 to 1981. η_j is the elasticity of interest, capturing the marginal change in patent output with respect to a change in firm f's real research-inputs during semi-decade j. γ_j are semi-decade fixed effects that capture average idea creation in a particular five-year period that is common to all firms in the sample in those years, after controlling for variation in their expenditure on research inputs.[25] Firm fixed effects γ_f ensure our estimates are identified from

[25] We estimate elasticities at the semi-decade frequency and use semi-decade fixed effects to minimize Census disclosure

variation within a firm over time.

We use each of the four measures of research inputs described in Section 3.4 – R&D expenditures, payroll in R&D labs, payroll in Professional and Management Services establishments, and total payroll – in four separate specifications. Although research inputs are a flow in the growth models, there may be a lag between inputs and outputs in our firm-level setting. In practice, we find that a depreciated 'stock' of the firms' past research-input expenditures captures a range of lags well, while reducing disclosure burden and enabling us to estimate time-varying elasticities.[26]

We estimate equation (8) via Poisson-Pseudo Maximum Likelihood (PPML). The set of firms included in each regression is restricted to those with at least one patent, citation, or breakthrough over the period and for which we observe an input $flow$ in that year, as we set stocks to missing if flows are not present. For the payroll-based stocks, this refinement is equivalent to dropping firms in any year in which they do not have those establishments. For the LBD R&D stock, firms are dropped when they do not appear in the RADS with positive R&D expenditures. Standard errors are clustered by firm.

5.2 Patent Elasticity Results

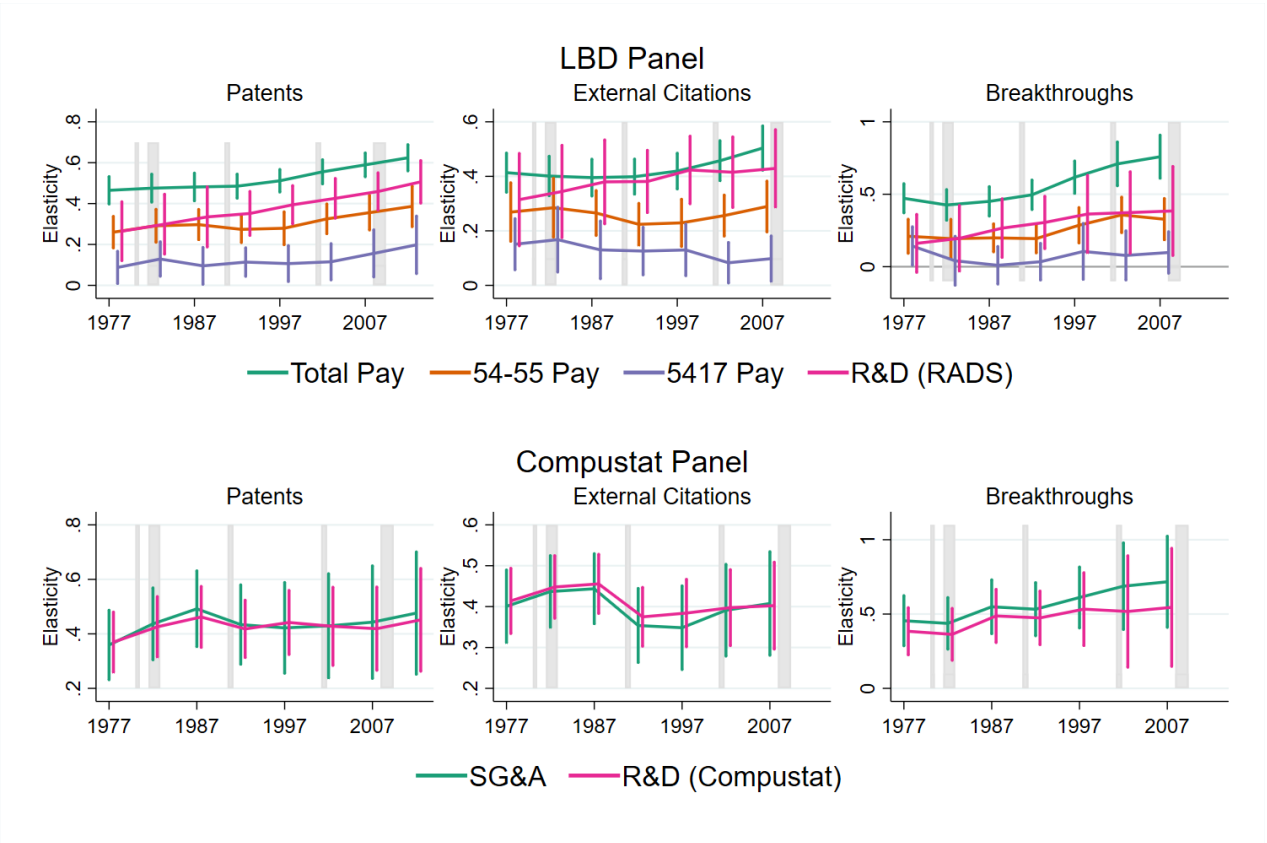
The top row of Figure 8 plots our estimates of η_j for the LBD panel. The elasticity of patents to R&D expenditure (upper left panel), which is most comparable to past work, ranges from 0.26 to 0.48 between the 1977 and 2012 semi-decades, indicating that a 1.0 log point increase in R&D expenditure is associated with 0.22 (=0.48-0.26) log point more patents at the end versus beginning of our sample period. These magnitudes are in line with those in Hausman et al. (1984), who estimate the elasticity of patents to R&D expenditure for 128 firms from 1968 to 1974 using a Poisson specification with firm fixed effects and a time trend. They find a same-year elasticity of 0.31 when lags for the previous 5 years are included, and 0.35 when lags are excluded (see their Table 2). However, when also including an interaction of the time trend with current R&D expenditures, their estimates imply a *decline* in the elasticity from 0.48 to 0.34 over their eight-year period. The sizes of our estimates are thus similar, though we find rising elasticities in this larger sample of firms and longer time horizon.

The middle and right panels in the top row of Figure 8 plot the estimated elasticities of external citations and breakthrough patents to R&D expenditure, which yield a similar message: the marginal effect of an added dollar of R&D by the end of the period is higher (though not necessarily statistically significantly so) than it was in the late 1970s. We find no evidence that the relationship between R&D inputs and ideas has weakened over the last 40 years. The remaining estimates in the top row of Figure 8 expand beyond the R&D survey sample by estimating the elasticity of all three patent measures to firms' Professional Services and Management pay, R&D lab pay, and total payroll. Across all but one of the measures of patenting and research inputs, elasticities are rising or flat over time. The exception is the elasticity of external citations to R&D lab pay. Since patent coverage by firms with R&D labs exhibits the most substantial decline across our measures of research input (see Figure

burden. Estimations employing year fixed effects yield analogous conclusions.

[26] We use a 15 percent depreciation rate. See Appendix Section E for details. In line with our findings, Hausman et al. (1984) show that the current flow of R&D expenditures generally subsume lags once firm fixed effects are included.

Figure 8: Estimated Patent Elasticities by Semi-Decade (η_j)



Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, DS, Compustat and authors' calculations. Notes: Figure reports estimates of η_j's from Equation 8 by patenting activity and research input for the LBD and Compustat panels. Whiskers denote 95 percent confidence intervals. Standard errors are clustered at the firm-level. Underlying coefficients and standard errors are reported in Appendix Tables A.15 and A.16.

7), we view these estimates with some caution.

Across the four research inputs, patent elasticities are generally highest for the total payroll stock and lowest for R&D lab pay. Some of these differences may arise from the very different levels in patenting across firm types. Firms in the total pay sample average 0.22 patents per year in the late 1970s, while R&D labs average 4.3 in that period (Appendix Table A.10). Another possibility is that R&D labs may increasingly provide services to other firms, whereas in the past they focused on in-house R&D. Such outsourcing will reduce our ability to link their R&D inputs to idea outputs. Consistent with this explanation, the R&D labs sample grows five-fold over the period, (Appendix Table A.11), while their average number of patents per firm falls almost in half (Appendix Table A.10). This changing sample composition for R&D labs is also evident in the firm fixed effect estimates, which we discuss in more detail below.

The top row of Figure 8 indicates that R&D inputs are at least as effective at producing ideas by the 2010s as they were in the late 1970s for the universe of firms. The second row of Figure 8 depicts the same message among Compustat firms. The elasticities of patents and breakthroughs to both R&D and SG&A expenditure are flat or rising. For external citations, they rise initially, fall during the early 1990s and then mostly recover by the end of the period.

These within-firm estimates provide direct evidence on the mapping between changes in a firm's R&D expenditure and its output of new patents (ideas), while controlling for the fact that more productive firms may spend more on R&D and be better at patenting. An important limitation of the estimates, however, is that they may be biased by time-varying firm-level shocks that simultaneously increase both R&D expenditures and patents. We note, however, that for such a bias to drive the generally rising estimates we document, it would need to be the case that the size of these firm-level shocks grows over time. While we cannot rule that out, if such shocks were the source of our results, that pattern in and of itself would seem to contradict the notion that ideas are getting harder to find.

5.3 Semi-Decade Fixed Effects Results

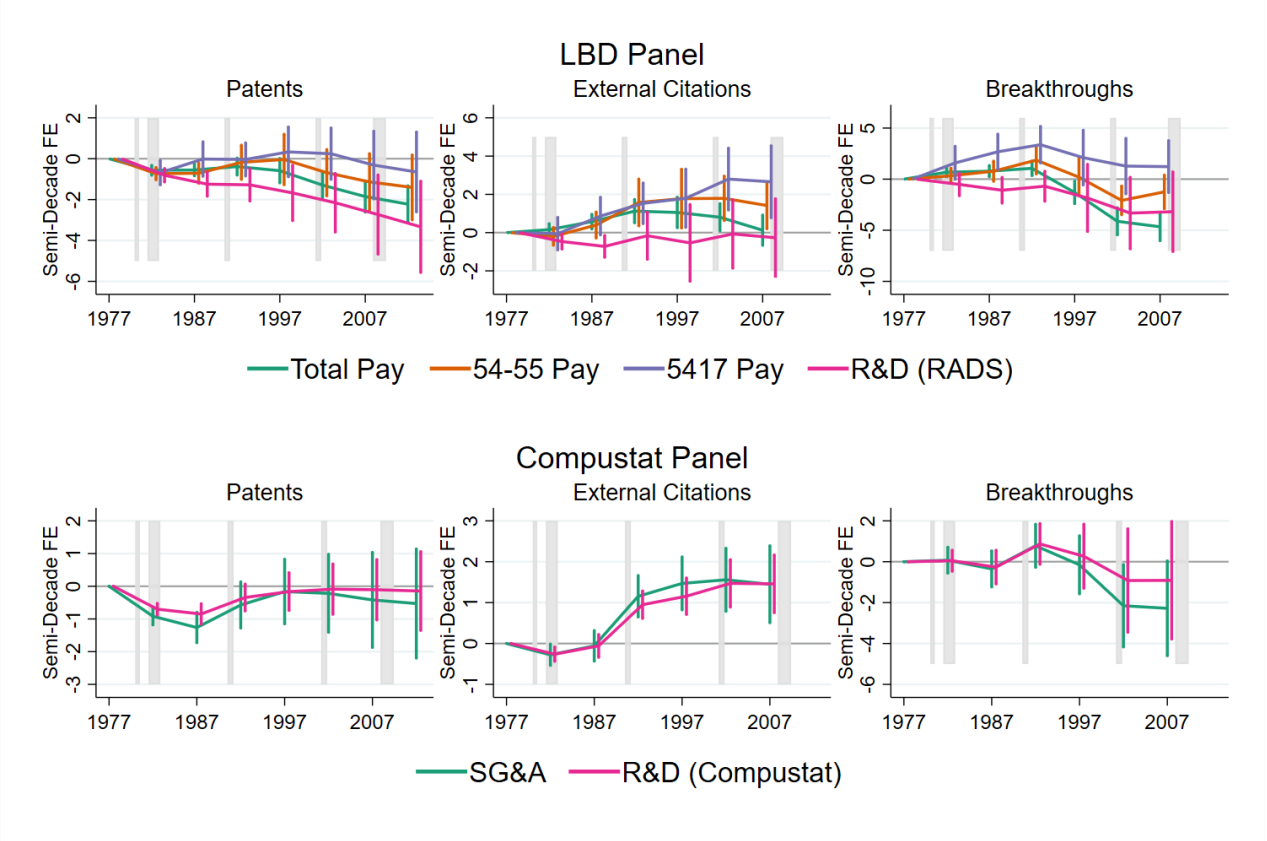
The results in Figure 8 suggest that it is *not* getting harder to generate ideas from research inputs. The elasticities of patents to research inputs are generally flat or rising over the last four decades, indicating that the marginal R&D dollar is more effective at generating a new idea at the end of the period than it was in the beginning. As discussed in Section 2, Romer (1990) predicts rising patents per researcher as A grows, whereas Jones (1995) predicts a flattening in researcher productivity gains whenever $0 > \beta > 1$. To the extent that a rising stock of knowledge makes idea generation relatively harder for all firms independent of their research inputs, this third concept of ideas getting harder to find is captured by the time fixed effects in equation (8). Given that patents have grown fairly steadily over our sample period (Figure 2), $\beta > 0$ would imply a secular flattening or decline in our estimated semi-decade fixed effects (assuming that idea obsolescence is not increasing). In practice, of course, the semi-decade fixed-effects also capture any other trends that are common across firms, such as changes in patenting laws or macroeconomic fluctuations.

We plot the estimated semi-decade fixed effects (γ_j) in Figure 9. These estimates capture average firm patents per period, after controlling for variation in R&D inputs, and thus provide a third way in which 'ideas getting harder to find' may manifest in the data. The first panel in the top row depicts a decline in these estimates when using firm R&D expenditure and total payroll as research inputs, but fairly flat coefficients when employing Professional and Management (NAICS 54-55) pay or R&D lab (NAICS 5417) pay. For external citations (second panel), the time effects are flat or rising across all research-input measures. For breakthroughs (third panel), estimates rise in the 1990s, decline until the early 2000s, and then flatten out.[27] The bottom row of Figure 9 reports results for the publicly traded firms in Compustat. Among these firms, there is no clear trend for patents and breakthroughs, and decidedly positive trends with respect to external citations.

Another way to illustrate the role of the semi-decade fixed effects is to show how they contribute to predicted versus actual US patenting. If patent elasticities and research inputs are both growing, then ideas should rise increasingly over time. On the other hand, if there is some aggregate factor reducing patenting for all firms, and it is more binding with time – as would be the case if a growing stock of knowledge crowds out new ideas – the time fixed effects would induce fewer predicted patents. To illustrate the role of the time fixed effects in the evolution of aggregate US patents, we predict each

[27] In additional, as-yet undisclosed robustness exercises, we find that this message is even stronger when we include firm size and age bins in the regressions.

Figure 9: Estimated Semi-Decade Fixed Effects from Patent Elasticity Regressions (γ_j)



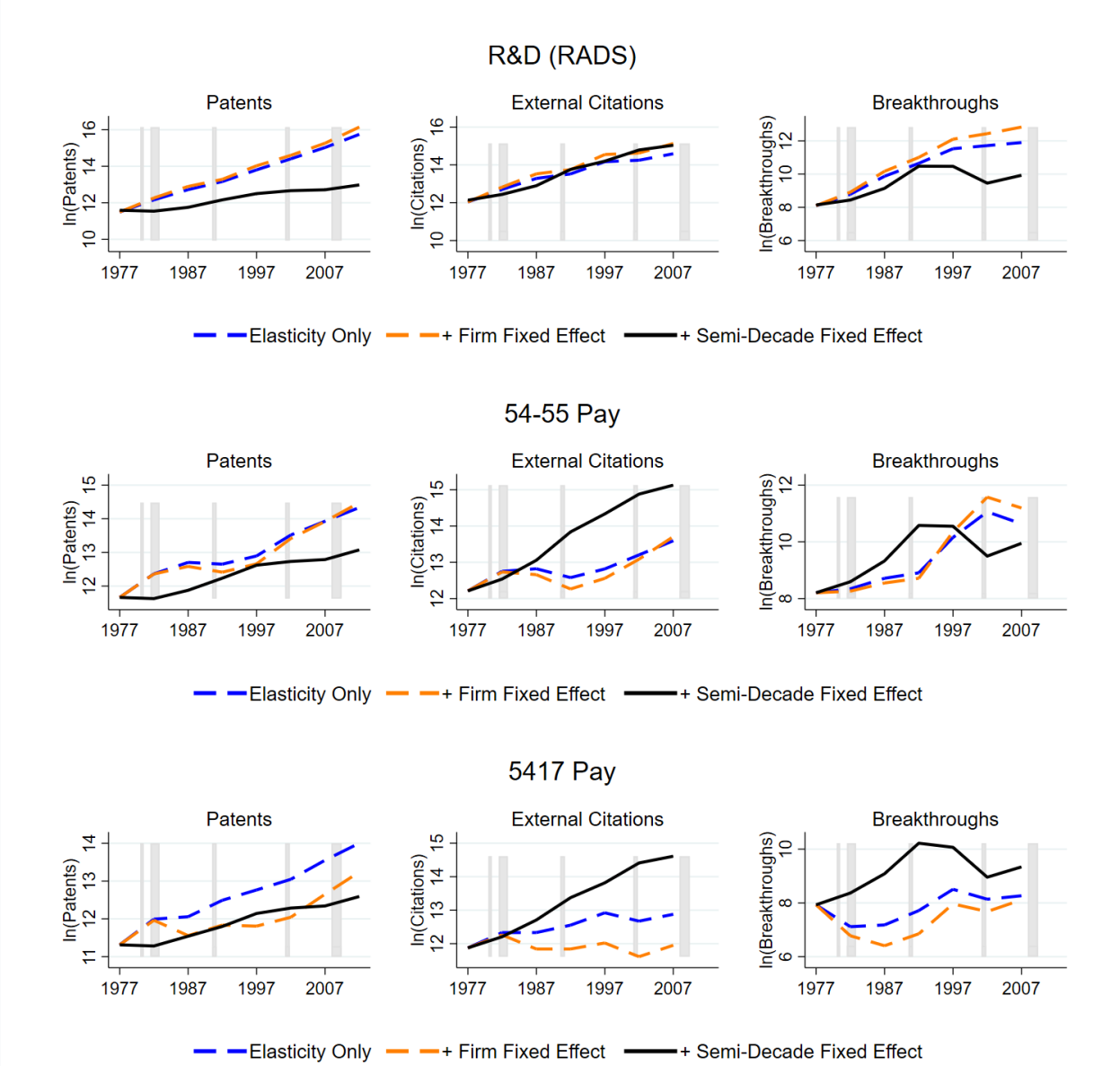
Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, Compustat and authors' calculations. *Notes:* Figure reports estimates of γ_j from Equation 8 by patenting activity and research-input stock for the LBD and Compustat panels. Whiskers denote 95 percent confidence intervals. Standard errors are clustered by firm. Underlying coefficients and standard errors are reported in Appendix Tables A.15 and A.16.

firm's patenting using our estimated coefficients and sum these predictions across firms. Although actual and predicted patents differ at the firm level, they match by construction in aggregate.

Figure 10 displays the results. The blue lines plot predicted patents using only firms' observed research-input use and the estimated patent elasticities. These generally rise with time. The orange lines add in the role of the firm fixed effects. While time-invariant for each firm, their contribution to overall predicted US can vary across the sample period as the composition of the sample varies. The final, black line adds in the contribution of the semi-decade fixed effects, matching actual US patents by construction. All lines are indexed to the actual number of US patents, external citations, and breakthroughs in 1977. As a result, each line provides information about how the contribution of each factor changes over time, but the lines cannot be compared in terms of levels.

The first panel in Figure 10 illustrates how the estimated semi-decade fixed effects in the first panel of in Figure 9 provide some support for the notion that granted *patents* are getting harder to find: while predicted patents based on firms' R&D expenditure rise throughout the period (blue dashed line), actual patent growth inclusive of the time effects is substantially more muted. External citations, however, reveal no such evidence, as the semi-decade fixed effects have either no effect (top row, middle panel) or a positive effect (rows 2 and 3, middle panels). For breakthroughs, the time

Figure 10: Predicted Patenting Activity Using Patent Regression Estimates (LBD Panel)



Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, and authors' calculations. Notes: Figure reports aggregate predicted patenting activity by semi-decade over the 1977 to 2007 (external citations, breakthroughs) or 2012 (patents) sample period. In each case, the aggregate prediction is the sum of predicted firm-level outcomes. Predictions are decomposed into the contribution of research inputs, that plus the contribution of the firm fixed effects, and that plus the contribution of the semi-decade fixed effects (see equation 8). The final line corresponds to the actual growth of each activity in our regression samples.

fixed effects contribute greater growth in the first half of the sample, then less in the second half. For firms with R&D labs, the firm fixed effects also play a role, as they increasingly pull down the number of predicted patents, consistent with changing composition of that sample to include more firms with much lower patents per firm (see Appendix Table A.10). Overall, we do not find systematic evidence that the average number of patents per firm declines over time when examining high-quality patent measures and a range of research inputs.

Intriguingly, the surge and subsequent decline in breakthrough patents midway through our sample period suggests that the emergence of whole new areas of innovation – in this case the PC/internet boom – can deliver a temporary burst of ideas. The subsequent leveling off of idea creation within such areas as they mature accords well with the micro evidence presented in Jones (2009), the sector-level analyses in Bloom et al. (2020), and the theoretical models of Evenson and Kislev (1976) and Ribeiro (2025), where applied research within a particular technology has diminishing returns over time, whereas basic research opens new frontiers. In more recent work, Teng (2025) shows that in the pharmaceutical industry, the declining rate of discovery of new functional molecular groups is offset by new drugs that combine such groups. In line with that work, we note that breakthroughs start growing again by the end of the period, perhaps as firms increasingly combine those IT inventions with other products as shown in Braguinsky et al. (2025).

Taking Stock To summarize the overall findings so far: Fact 1 shows that research productivity, measured as total patents over R&D inputs, has risen for the universe of firms in the United States, even as it declined among Compustat firms. Figure 8 demonstrates that the elasticity of patent activity to R&D inputs is less than one, but flat or rising over time. Finally, Figure 9 does not depict a consistent secular flattening or decline in average firm patent activity across our various measures as the aggregate stock of knowledge grows. Indeed, the only data in which we observe some support for the premise that ideas are getting harder to find is for patent grants with respect to R&D expenditure. The preponderance of the evidence therefore does not support the premise that a growing stock of knowledge crowds out future ideas, *i.e.*, ideas are *not* getting harder to find. On the other hand, Romer (1990) predicts that researcher productivity should rise as the aggregate stock of ideas grows. While we estimate rising patent elasticities for some measures, they are relatively flat for others.

The results so far only provide a partial explanation for the divergence in TFP and research inputs depicted in Figure 1. The estimated elasticities of patents to R&D inputs are less than one, which indicates that output growth and research inputs will diverge over time, though the former should nevertheless increase as the latter grows. Moreover, the rising elasticities should lessen the divergence over our period. If the mapping from ideas to output is one-for-one, as derived in Romer and other models, increasing ideas should map to rising output growth.[28] We turn to estimating that relationship directly in the next section.

6 Estimating the Link Between Ideas and Growth

If ideas are not getting harder to find, why has aggregate growth not kept pace with rising researcher efforts? A key assumption in many macro-growth models is a one-to-one mapping between growth in ideas and growth in output that is constant over time ($\dot{A}/A = \dot{Y}/Y$). In this section, we estimate the relationship between ideas and growth directly to assess empirical support for this assumption.

[28] An important caveat is that the Romer null hypothesis assumes proportional growth in ideas from a fixed flow of R&D inputs. The flat or rising elasticities we estimate indicate that the level of patents grows with increased R&D, but does not guarantee a higher *growth rate* of ideas from higher R&D expenditure.

6.1 Aggregate Growth in the LBD Panel

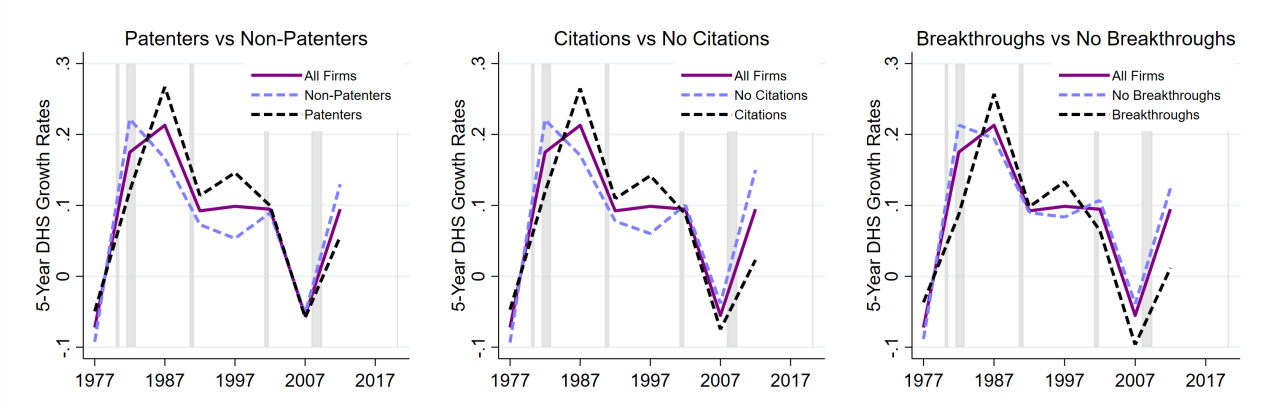
Before proceeding with formal estimates, we use our LBD panel to quantify the role of patenting firms in overall US sales growth. The three panels of Figure 11 compare growth rates of firms with at least one patent, external citation, or breakthrough to all other firms. Because the EC data are only available every five years, each point corresponds to the growth rate from t to $t+5$. For example, the 1977 point depicts growth from 1977 to 1982.

The first message of this figure is that patenting firms have relatively high sales growth from 1987 through 2002. Despite comprising less than one percent of all firms, their growth is sufficiently high and their size sufficiently large to pull up aggregate growth rates during that period, relative to what they would have been if only the non-patenters existed. Indeed, from the 1987 to 2002 period, patenting firms account for 58 to 72 percent of total US private, non-farm employer growth.

The second message from Figure 11 is that patenting firms' growth premia disappear after 2002. In fact, by the 2010 decade, patenting firms grow relatively less than non-patenters. This pattern raises the possibility that the relationship between ideas and growth might be weakening towards the end of the sample period.

The spike in aggregate growth rates during the 1987 to 1992 semi-decade coincides with the addition of the Economic Census of Finance, Insurance, and Real Estate in 1992. In undisclosed results, we drop establishments in year $t+5$ if they are in the LBD but missing from an any of the Economic Censuses in year t, and vice-versa. Increasingly stringent disclosure avoidance standards limit our ability to include those results in the current draft, but we expect to include them in future versions of this paper.[29]

Figure 11: Aggregate Sales Growth By Patenting Activity



Source: LBD, EC, PV, KPST, and authors' calculations. Notes: Figure reports the aggregate sales growth from t to t_5 across all firms, and by their patenting activity, *i.e.*, those with at least one patent grant, external citation, or breakthrough over the entire period, versus not. The black dashed lines in this figure are equivalent to those in Figures 12 and 15. The value for 1977 represents growth from 1977 to 1982, etc.

[29]The Census Bureau imposes limits on the number of estimates that can be disclosed for a particular sample and analysis. We are waiting to disclose additional results in response to future comments and suggestions to retain as much disclosure flexibility as possible.

To understand which firm margins drive the patterns above, we decompose the total sales growth of patenting and non-patenting firms into the contributions from entrants, exiters, and continuers, defining growth as

$$Growth\left(Sales_f^{t:t+5}\right) = \frac{Sales_f^{t+5} - Sales_f^t}{\frac{1}{2}\left(Sales_f^{t+5} + Sales_f^t\right)}. \tag{9}$$

This growth rate is identical to a log change for continuers, and equal to 2 and -2 for entrants and exiters. The three dashed lines in each panel of Figure 12 plot the contribution from each of these margins to total growth for patenting firms (top panel) and non-patenting firms (bottom panel). Within each panel, the sum of each margin (i.e., the dashed lines) is equal to total growth by that patenting firm type (i.e., the black line) by construction. The solid black lines in each panel of Figure 12 thus match the total growth rates by patenting and non-patenting firms depicted in Figure 11.

Figure 12 clearly shows that continuing firms account for the bulk of patenting firms' growth, particularly during the 1997 to 2002 semi-decade. For non-patenters, by contrast, entrants are most salient. Also noteworthy in Figure 12 is the secular decline in net job creation by non-patenting firm entrants and continuers. This decline coincides with smaller negative contributions from exiters. The decrease in exits, along with falling net job creation from entrants, aligns with the decline in churn in US employment dynamics documented in past work (Decker et al., 2020).

Overall, the trends in Figure 11 indicate that patenting firms raised aggregate growth during the 1987 to 2002 period. By contrast, patenters grew relatively less than non-patenters in later periods. Figure 12 shows that continuing firms play the dominant role in patenting firms' growth, although reduced entry and higher exit explain part of their slower growth in the 2000s and 2010s. We now turn to formal estimation of patents' role in patenting firms' growth.

6.2 Growth Regression Specification

We analyze the relationship between firms' sales growth and their new ideas by adapting equation (7) to estimate:

$$\begin{aligned} Growth\left(X_f^{t:t+5}\right) = &\sum_{j=1977(5)k} \mu_j d[j]_{ft} \times Growth\left(IdeaStock_f^{t:t+5}\right) + \sum_{j=1977(5)k} \phi_j d[j]_{ft} \\ &+ \sum_{b=1:5} \rho_b^S \{SizeBin_{fb}\} + \sum_{b=1:5} \rho_b^A \{AgeBin_{fb}\} + \nu_{ft} \end{aligned} \tag{10}$$

where the left-hand side is the growth rate of firm f's sales per worker between t and t+5 and growth is defined as in equation 9 above. μ_j is the coefficient of interest, capturing the influence of new ideas on growth. $Growth\left(IdeaStock_f^{t:t+5}\right)$ is the growth rate in a firm's number of patents, external citations, or breakthroughs in each semi decade. The final two terms in equation 10 are indicators for firm size and age bins, which we use as proxies for capital. While Census data have an explicit measure of capital for manufacturing firms, Fact 2 demonstrates that a complete analysis US patenting over

Figure 12: Aggregate Sales Growth Contributions By Patenting Activity and Margin



Source: LBD, EC, PV, KPST, and authors' calculations. Notes: Figure reports the aggregate sales growth from t to $t+5$ across firms by their patenting activity and margin. The value for 1977 represents growth from 1977 to 1982, etc.

.

our sample period must extend beyond that sector.[30] These controls also address potential variation in average firm growth rates arising from changes in the composition of firms over time (Fact 3). As in our earlier regressions, ϕ_j represent semi-decade fixed effects. Their estimated coefficients capture secular trends in sales per worker growth driven by the other factors that are not included in the specification.

We estimate equation (10) via Ordinary Least Squares on the subset of continuing firms in each of the patent samples (*i.e.*, the dashed blue line in Figure 12). We also estimate a variant of equation (10) with firm sales as the dependent variable. That specification includes entrants and exiters as well as continuers, and includes an additional control for firm employment growth, $Growth\left(Employment_f^{t:t+5}\right)$. We follow Davis et al. (1996) and weight the regressions by firms' average sales in t and $t+5$, *i.e.*, the denominator in equation (9). This weighting allows us to recover the sample aggregates.

[30] We use the same firm size and age bins depicted in Figure 5.

6.3 Growth Regression Results

Figure 13 presents the estimated coefficients from equation (10) with growth in firms' sales per worker as the dependent variable.[31] The top panel depicts estimates for the coefficients on firms' ideas (μ_j). For both patent grants and citations, we estimate a positive and statistically significant relationship between the growth of new ideas and sales per worker in every time period. Although growth in breakthrough patents is unrelated to productivity growth in the first semi-decade (1977 to 1982), a 10 percentage point increase in the growth rate of breakthrough patents is associated with a 0.03 percentage point increase in the firm's productivity growth in the 2007 to 2012 period.

Most notably for our question, the responsiveness of a firm's productivity growth to growth in its ideas is larger at the end of the period for all three measures of ideas. For patent grants, a 10 percentage point increase in the firm's growth rate of patents is associated with a 0.012 percentage point increase in its sales per worker growth rate in 1977, versus a 0.037 percentage point increase in the 2012 to 2017 semi-decade. Overall, the estimates suggest a strong, positive relationship between firm productivity improvements and new ideas that strengthens over time.

The bottom panel of Figure 13 reports the semi-decade fixed effects from the same specifications. The 1977 to 1982 period is the omitted category, so each estimate shows average firm growth over the period relative to growth in the initial period, after controlling for changes in a firm's ideas, as well as its size and age. Across the three specifications and samples, we estimate rising residual growth rates until the early 1990s, after which there is a fairly steady decline. By the end of our sample period, firms' average labor productivity growth rates are close to their 1977 levels (and not statistically different from them).[32]

We also estimate equation (10) using the growth rate of firm sales as the dependent variable. In this specification, we control for firm employment growth, $Growth\left(Emp_f^{t:t+5}\right)$, to account for increases in sales due to factors other than innovation. To the extent, however, that new ideas lead firms to increase employment, we will underestimate the impact of ideas on growth. Unlike the specifications with growth in labor productivity as the dependent variable, we can include entrants and exiters in these specifications. The sample for these regressions thus aligns with the patenting firm specifications in Section 5.1.

Figure 14 presents these results. Firm sales growth is positively associated with patent grant and citation growth in all semi-decades, and is statistically significant in most of them. The estimates for breakthrough patents follow an analogous pattern, though they are noisier and only statistically significant in the 1997 (1997 to 2002) and 2002 (2002 to 2007) semi-decades. The semi-decade fixed effects in these regressions follow a similar trajectory to those in the sales per worker specifications (Figure 13): they rise after the first 1977 to 1982 period, are fairly flat through the 2002 to 2007 period, and then decline.[33] A notable difference between the sales per worker versus sales regressions

[31] Appendix Table A.17 reports the estimates.

[32] As noted above, there is a spike in growth rates and thus also in these semi-decade fixed effects for 1987 to 1992, some of which is due to the fact that the Census of Finance is first available in 1992. In undisclosed results, we drop establishments in year $t+5$ if they are in the LBD but missing from an any of the Economic Censuses in year t, and vice-versa.

[33] Again, here we ignore the spike for the 1987 to 1992 period.

Figure 13: Estimated Sales/Worker Growth Coefficients (LBD Panel)



Source: LBD, EC, PV, KPST, and authors' calculations. Notes: Figure reports relationship between LBD firms' sales per worker and idea stock growth between years t and t+5 Standard errors are clustered at the firm-level with 95 percent confidence intervals depicted. Appendix Table A.17 reports corresponding estimates.

is the latter's coefficients on patent growth are substantially smaller, both compared to the estimated patent growth coefficients and the semi-decade fixed effects. This smaller magnitude may, at least in part, reflect our inclusion of firm employment growth as an additional control variable that co-moves with ideas in driving sales.

The key message of Figures 13 and 14 is that firms' new ideas are associated with higher growth rates, and this association is flat or rising over time. On the other hand, secular trends in firms' average growth rates – captured by the semi-decade fixed effects – work against the contribution of ideas, thus lowering patenting firms' aggregate growth in later years back to their 1977 to 1982 levels. These secular declines echo the declining growth-rate contributions of continuing and entering non-patenting firms depicted in the bottom row of Figure 12. In the specifications with sales per worker as the dependent variable, the estimated coefficients on ideas and semi-decade fixed effects are of similar magnitude. In the specifications with sales as the dependent variable, where the specification also controls for growth in employment, μ_j is substantially smaller than the estimated semi-decade effects, which in turn implies a much smaller role for ideas in predicting growth. In Appendix Section I.1, we show that these patterns are also evident in unweighted regressions.

Figure 14: Estimated Sales Growth Coefficients (LBD Panel)



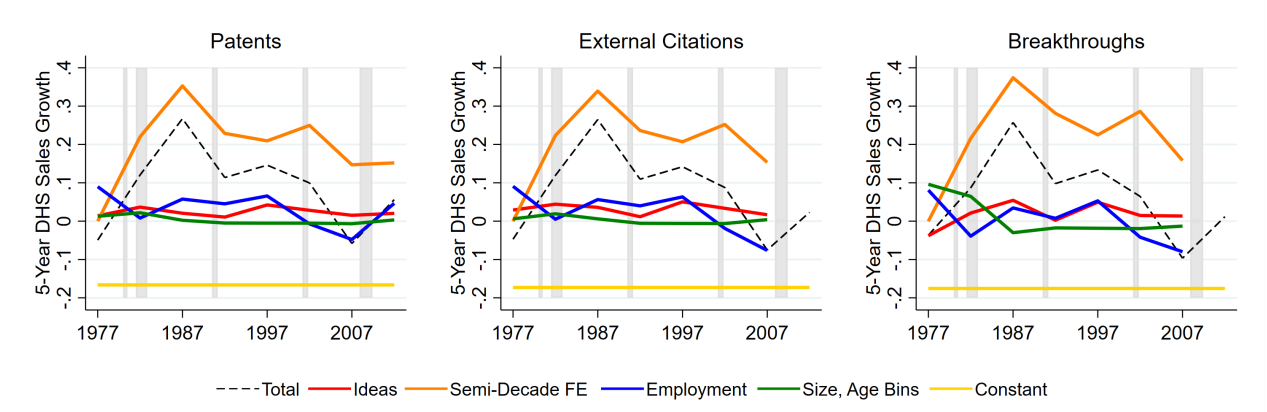
Source: LBD, EC, PV, KPST, and authors' calculations. Notes: Figure reports relationship between LBD firms' sales and idea stock growth between years t and t+5 Standard errors are clustered at the firm-level with 95% CIs depicted. Appendix Table A.17 reports corresponding estimates.

To illustrate the contribution of ideas versus other forces to the aggregate sales growth depicted in Figure 11, we again employ our coefficient estimates to predict aggregate sales growth attributable to ideas, size, age, employment and the semi-decade fixed effects. Along with the constant, the sum of these contributions exactly matches patenting firms' aggregate growth by construction. We focus on the specifications with sales growth as the outcome variable because these regressions include entrants and exiters and thus exactly match the dashed black line for patenters' total growth rates in Figure 11.

Results are displayed in Figure 15. The red lines show predicted sales growth from idea creation, *i.e.,* the summation across all firms of the product of μ_j and firms' idea growth. For patents (left panel) this contribution is positive but small in all periods, with a marked increase for the 1997 to 2002 semi-decade. This predicted increase aligns with the time period in which patenting firms grow relatively more than non-patenters (Figure 11), suggesting that new ideas played a key role in their disproportionate performance. The contribution of employment (yellow line) is negative and sizable starting after the 2002 to 2007 semi-decade, indicating that firms reduced employment growth considerably in those years, and sales growth followed suit. This period corresponds to the

disappearance of patenters' growth premia (Figure 11), indicating that the impact of ideas on growth may have been swamped by macroeconomic factors leading firms to downsize. Firms' size and age (blue lines) play a fairly small role explaining aggregate sales growth that is lower at the end of the sample than the beginning. This trend is consistent with a shift of economic activity towards large and old firms which have comparatively lower growth rates (Haltiwanger et al., 2013). Finally, the semi-decade fixed effects (orange lines) predict substantial increases in sales growth after the initial 1977 to 1982 period, but decline in later years.[34]

Figure 15: Contributions of Firm Sales Growth Regression Terms to Aggregate Sales Growth



Source: LBD, EC, PV,KPST, and authors' calculations. Notes: Figure reports the contribution of each term in the sales growth regression (equation 10) to predicted growth. The dashed black line tracks aggregate sales growth across the firms in each sample. The solid colored lines indicate the contributions of: idea growth (red), the semi-decade fixed effects (orange), employment growth (gold), the constant (green), and the age and size bins (blue), all of which sum to the total (black).

To gain a sense of the relative magnitudes, the first panel in Table 1 reports total growth in each semi-decade, along with the predicted growth from firms' actual idea and employment growth for the patent grant sample. The residual ("Other") reports predicted net growth due to the time, size, and age fixed effects. Predicted growth from ideas for each five-year period ranges from 1 to 4 percent. It is highest (4 percent) in the 1982 to 1987 and 1997 to 2002 periods. Although it is lower in the 2007 to 2017 decade (2 percent), the first half of that period coincides with negative total growth as both employment and the time fixed effects pick up shrinking output during the Great Recession. If anything, ideas seem to have sustained growth during this period, even as other macro factors pulled it down. Most notably, we do not observe steadily declining predicted growth rates from ideas, as would be predicted by semi-endogenous growth models if the concavity in the idea production function were starting to bite by reducing the impact of ideas on output. On the other hand, we also do not observe steadily rising predicted growth rates from ideas that would suggest Romer's compounding forces are at play.[35]

[34] As noted above, a portion of the spike in these semi-decade fixed effects for 1987 to 1992 is due to the fact that the Census of Finance is first available in 1992.

[35] In ongoing work, we are calculating predicted growth from firm's *predicted* ideas to assess whether and how much

Table 1: Predicted Growth from Ideas vs Other Factors

	Patent Grant Sample					Citation Sample		Breakthrough Sample	
	Total	Predicted Growth From:			Grant	Predicted Growth From:		Predicted Growth From:	
	Growth	Grants	Emp	Other	Share	Citations	Share	Breakthroughs	Share
1977-1982	-0.05	0.01	0.09	-0.15	-0.20	0.03	-0.60	-0.04	1.00
1982-1987	0.12	0.04	0.01	0.08	0.33	0.04	0.33	0.02	0.22
1987-1992	0.27	0.02	0.06	0.19	0.07	0.04	0.15	0.05	0.19
1992-1997	0.11	0.01	0.05	0.06	0.09	0.01	0.09	0.00	0.00
1997-2002	0.15	0.04	0.07	0.04	0.27	0.05	0.36	0.05	0.38
2002-2007	0.10	0.03	-0.01	0.08	0.30	0.03	0.33	0.01	0.17
2007-2012	-0.06	0.02	-0.05	-0.03	-0.33	0.02	-0.25	0.01	-0.10
2012-2017	0.06	0.02	0.05	-0.01	0.33				
1977-2017	0.70	0.19	0.26	0.25	0.27	0.22	0.35	0.12	0.23

Source: PV, EC, LBD, KPSS, KPST, BDSPF-Long, RADS, and authors' calculations. Table reports total growth by period, as well as predicted growth using the estimated coefficients depicted in Figure 14 along with firms' actual patent growth, employment growth, and the time and size-age bins. Grant Share is the fraction of total growth over the period that is predicted from firms' actual growth in patent grants. Citation and breakthrough samples report the contributions of citation and breakthrough growth only, and their shares of total growth in those samples.

The next two panels in Table 1 report the predicted growth of citation and breakthrough growth, along with the share of the total growth in those samples that this predicted growth comprises. Predicted growth from citations is generally the same or higher than predicted growth from patent grants. For breakthroughs, predicted growth is more volatile, with two high points in the 1982 to 1992 and 1997 to 2002 periods. Across all measures, patent growth predicts the highest sales growth during the 1997 to 2002 period, which corresponds to a period with particularly high elasticity estimates and patent growth. Predicted growth is lower after this burst, and especially for breakthroughs. This decline is not surprising given the inverted-U shaped pattern of breakthrough patents. Indeed, the volatile patterns for breakthroughs highlight the unpredictable nature of innovation and the fact that it grows in sharp bursts rather than at a constant rate.

Figure 15 and Table 1 make clear that while a firm's ideas are positively related to its sales growth, their impact varies over time and is sometimes swamped by other macro factors. Some of these factors affect all firms equally (the time fixed effects), while others (employment) are firm-specific. The latter are especially low during the 2007 to 2012 period, which aligns with the Great Recession and is therefore not surprising. Indeed, these trends also correlate with the patterns among $non-patenting$ firms depicted in Figure 12.

One potential rationale for the secular decline in sales growth captured by the semi-decade fixed effects is that idea spillovers *across* firms are falling, even as firms' own ideas continue to generate growth. Such a reduction might also explain at least some of the declines in non-patenting firms' growth rates evident in the bottom panel of Figure 12. Of course there are many other potential explanations for anemic growth *after* controlling for changes in firms' ideas, *e.g.,* institutions, climate,

the declining year fixed effects in the patenting regressions affect the aggregate predicted growth from ideas we calculate here.

etc. Most importantly for our purposes, we find a positive and significant relationship between a firm's idea and output growth that does not fall systematically over time.

7 Conclusion

We build a novel 45-year firm-level panel to study the evolution of US patenting. In contrast to the view that ideas are getting harder to find, we document rising patents per R&D input, flat or rising elasticities of patents to R&D inputs, no robust secular decline in average patents per firm, and no weakening of the relationship between firms' patent and output growth. These results undermine the conclusion that flat US TFP growth coupled with rising research expenditure necessarily indicates declining researcher productivity. At least over the last four decades, any crowding out of ideas making future innovation harder does not seem to bite in the aggregate.

While the marginal R&D dollar today is at least as effective at generating ideas as it was 40 years ago, firms' average growth rates after controlling for their patent growth fall in the later periods. A natural next question is what has led to a decline in average growth rates across firms that is independent of their own idea creation? One possibility is that spillovers across firms have declined. If, for example, it is easier for firms to exclude rivals from their ideas, then average growth may decrease, even as ideas expand. Indeed, recent work highlights precisely how important declining knowledge diffusion seems to be for the US economy after 1980 (Akcigit and Ates, 2023).

We view a micro-level analysis of changes in such spillovers as the natural next step to unpacking the relationship between ideas and growth. The US increasingly specializes in the innovation stages of the production process, while downstream production stages and profits from the full value chain are often performed and accrued outside the United States (Fort, 2023). In line with these trends, Fact 2 shows that US patents have shifted towards Information, Professional Services, and Managements sectors, which generally provide inputs to downstream stages of production that may be classified in other sectors or located in other countries.[36] A fruitful avenue for future work is thus to trace out the cross-sectoral and cross-country patterns in which innovation in one area feeds into increased productivity in others and how such spillovers have evolved over time.

[36]For example, Braguinsky et al. (2025) document new product introductions in seemingly low-tech sectors such as apparel when firms such as Nike add GPS tracking to their clothing.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 2001, *91* (5), 1369–1401.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, "Competition and Innovation: An Inverted-U Relationship," *Quarterly Journal of Economics*, 2005, *120* (2), 701–728.

Aghion, Phillippe and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–351.

Akcigit, Ufuk and Sina T. Ates, "What Happened to U.S. Business Dynamism?," *Journal of Political Economy*, 2023, *131* (10), 2319–2369.

_ **and William R Kerr**, "Growth Through Heterogeneous Innovations," *Journal of Political Economy*, 2018, *126* (4), 1374–1443.

Ando, Yoshiki, James Bessen, and Xiupeng Wang, "The Rising Returns to RD: Ideas are not getting harder to find," Technical Report 5242171, SSRN 2025.

Aw, Bee Yan, Mark J. Roberts, and Daniel Yi Xu, "R&D Investment, Exporting, and Productivity Dynamics," *American Economic Review*, 2011, *101* (4), 1312–1344.

Balasubramanian, Natarajan and Jagadeesh Sivadasan, "What Happens When Firms Patent? New Evidence from US Economic Census Data," *Review of Economics and Statistics*, May 2011, *93* (1), 126–146.

Bergeaud, Antonin, Arthur Guillouzouic, Emeric Henry, and Clément Malgouyres, "From Public Labs to Private Firms: Magnitude and Channels of R&D Spillovers," *Quarterly Journal of Economics*, forth.

Bernard, Andrew B. and Teresa C. Fort, "Factoryless Goods Producing Firms," *American Economic Review*, May 2015, *105* (5), 518–23.

Bessen, James, "The Value of US Patents by Owner and Patent Characteristics," *Research Policy*, 2008, *37* (5), 932–945.

Bilir, L. Kamran and Eduardo Morales, "Innovation in the Global Firm," *Journal of Political Economy*, 2020, *128* (4), 1566–1625.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," *American Economic Review*, April 2020, *110* (4), 1104–44.

_ **, Mark Schankerman, and John Van Reenen**, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, July 2013, *81* (4), 1347–1393.

Blundell, Richard, Rachel Griffith, and Frank Windmeijer, "Individual Effects and Dynamics in Count Data Models," *The Quarterly Journal of Economics*, 2002, *117* (2), 481–530.

Braguinsky, Serguey, Joonkyu Choi, Yeheng Ding, Karam Jo, and Seula Kim, "Mega Firms and New Technological Trajectories in the US," Technical Report 31460, NBER 2025.

Brynjolfsson, Erik and Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York: W. W. Norton & Company, 2014.

Chen, Yanyou and Daniel Xu, "A Structural Empirical Model of R&D, Firm Heterogeneity, and Industry Evolution," *The Journal of Industrial Economics*, 2023, *LXXI* (2), 323–353.

Chow, Melissa C, Teresa C Fort, Christopher Goetz, Nathan Goldschlag, James Lawrence, Elisabeth Ruth Perlman, Martha Stinson, and T Kirk White, "Redesigning the Longitudinal Business Database," Technical Report 28839, National Bureau of Economic Research Working Paper Nr 28839 2021.

Cohen, Wesley M, Richard Nelson, and John P Walsh, "Protecting their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (Or Not)," Working Paper 7552, NBER 2000.

Cowen, Tyler, *The great stagnation: How America ate all the low-hanging fruit of modern history, got sick, and will (eventually) feel better: A Penguin eSpecial from Dutton*, Penguin, 2011.

Davis, Steven J, John Haltiwanger, and Scott Schuh, "Small business and job creation: Dissecting the myth and reassessing the facts," *Small business economics*, 1996, *8*, 297–315.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda, "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness," *American Economic Review*, 2020, *110* (12), 3952–3990.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken, "Temperature Shocks and Economic Growth: Evidence from the Last Half Century," *American Economic Journal: Macroeconomics*, 2012, *4* (3), 66–95.

Ding, Xiang, Teresa C Fort, Stephen J Redding, and Peter K Schott, "Structural Change Within versus Across Firms: Evidence from the United States," Working Paper 30127, National Bureau of Economic Research June 2022.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "The Regulation of Entry," *The Quarterly Journal of Economics*, 2002, *117* (1), 1–37.

Doidge, Craig, G. Andrew Karolyi, and René M. Stulz, "The U.S. Listing Gap," *Journal of Financial Economics*, 2017, *123* (3), 464–487.

Dreisigmeyer, David, Nathan Goldschlag, Marina Krylova, Wei Ouyang, and Elisibeth Perlman, "Building a Better Bridge: Improving Patent Assignee-Firm Links," Technical Report, Center for Economic Studies CES-TN-2018-01 2018.

Dyevre, Arnaud and Oliver Seager, "Matching Patents to Publicly Listed Firms in the US: 1950-2020," Technical Report, LSE 2024.

Ekerdt, Lorenz K.F., "The Role of R&D Factors in Economic Growth," CES Working paper 24-69, Center for Economic Studies 2024.

_ **and Kai-Jie Wu**, "Self-Selection and the Diminishing Returns to Research," Working paper, Center for Economic Studies 2024.

Evenson, Robert, "International Invention: Implications for Technology Market Ananlysis," in Zvi Griliches, ed., *R&D, Patents, and Productivity*, University of Chicago Press, 1984, pp. 89–126.

Evenson, Robert E. and Yoav Kislev, "A Stochastic Model of Applied Research," *Journal of Political Economy*, 1976, *84* (2), 265–282.

Fort, Teresa C., "The Changing Firm and Country Boundaries of US Manufacturers in Global Value Chains," *Journal of Economic Perspectives*, 2023, *37* (3), 31–58.

Fort, Teresa C and Shawn D Klimek, "The Effects of Industry Classification Changes on US Employment Composition," *CES Workign Paper*, 2018, (18-23).

Gordon, Robert, *The rise and fall of American growth: The US standard of living since the civil war*, Princeton university press, 2017.

Graham, Stuart JH, Cheryl Grim, Tariqul Islam, Alan C Marco, and Javier Miranda, "Business Dynamics of Innovating Firms: Linking US Patents with Administrative Data on Workers and Firms," *Journal of Economics & Management Strategy*, 2018, *27* (3), 372–402.

Griliches, Zvi, "Issues in Assessing the Contribution of Research and Development to Productivity Growth," *The Bell Journal of Economics*, 1979, pp. 92–116.

_ , "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 1990, *28* (4), 1661–1707.

_ , "Patent Statistics as Economic Indicators: a Survey," in "R&D and Productivity: the Econometric Evidence," University of Chicago Press, 1998, pp. 287–343.

Grossman, Gene M. and Elhanan Helpman, "Quality Laddres in the Theory of Growth," *Review of Economic Studies*, 1991, *58*, 43–61.

Growiec, Jakub, Peter McAdam, and Jakub Mućk, "Are Ideas Really Getting Harder To Find? R&D Capital and the Idea Production Function," 2022.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, "Market Value and Patent Citations," *RAND Journal of Economics*, 2005, *36* (1), 16–38.

_ , Jacques Mairesse, and Pierre Mohnen, "Measuring the Returns to R&D," in "Handbook in Economics," Vol. 2, Elsevier B.V., 1998, chapter 24, pp. 1034–1082.

Haltiwanger, John C., Ron S. Jarmin, and Javier Miranda, "Who Creates Jobs? Small vs. Large vs. Young," *Review of Economics and Statistics*, 2013, *95* (2), 347–361.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches, "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 1984, *52* (4), 909–938.

Howell, Sabrina T., "Financing Innovation: Evidence from R&D Grants," *Americal Economic Review*, 2017, *107* (4), 1136–1164.

Jaffe, Adam B. and Gaétan de Rassenfosse, "Patent Citation Data in Social Science Research: Overview and Best Practices," *Journal of the Association for Information Science and Technology*, 2017, *68* (6), 1360–1374.

Jarmin, Ron S and Javier Miranda, "The Longitudinal Business Database," CES Working Paper CES-WP-02-17 2002.

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

Jones, Chad I., "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 1995, *103* (4), 759–784.

Kamal, Fariha, "A Portrait of US Factoryless Goods Producers," in Nadim Ahmad, Brent Moulton, J. David Richardson, and Peter Van De Ven, eds., *Measuring the Global Economy*, University of Chicago Press, 2023.

_ , Jessica McCloskey, and Wei Ouyang, "Multinational Firms in the U.S. Economy: Insights from Newly Integrated Microdata," Working Paper 22-39, Center for Economic Studies 2022.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring Technological Innovation over the Long Run," *American Economic Review: Insights*, September 2021, *3* (3), 303–20.

Kerr, William R and Shihe Fu, "The survey of industrial R&D—patent database link project," *The Journal of Technology Transfer*, 2008, *33*, 173–186.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar, "Who Profits from Patents? Rent-Sharing at Innovative Firms," *The Quarterly Journal of Economics*, 2019, *132* (2), 1343.1404.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological Innovation, Resource Allocation, and Growth*," *The Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

Koh, Ping-Sheng and David M. Reeb, "Missing R&D," *Journal of Accounting and Economics*, 2015, *60*, 73–94.

Kortum, Samuel, "Equilibrium R&D and the Decline in the Patent-R&D Ratio: U.S. Evidence," *American Economic Review Papers and Proceedings*, 1993, *83*, 450–457.

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

Lanjouw, Jean O. and Mark Schankerman, "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators," *Economic Journal*, 2004, *114* (495), 441–465.

Meyers, Kyle R. and Lauren Lanahan, "Estimating Spillovers from Publicly Funded R&D: Evidence from the US Department of Energy," *American Economic Review*, 2022, *112* (7), 2393–2423.

Peters, Bettina, Mark J. Roberts, Van Anh Vuong, and Helmut Fryges, "Estimating Dynamic R&D Choice: An Analysis of Costs and Long-Run Benefits," *RAND Journal of Economics*, 2017, *48* (2), 400–437.

Philippon, Thomas, "Additive Growth," NBER Working Paper 29950, National Bureau of Economic Research 2022.

Ribeiro, Bernardo, "Growth with New and Old Technologies," Working Paper, Princeton University 2025.

Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98* (5), S71–S102.

Schlingemann, Frederik P. and René M. Stulz, "Have Exchange-Listed Firms Become Less Important for the Economy?," *Journal of Financial Economics*, 2022, *143* (2), 927–958.

Teng, Songyuan, "Innovation Through Recombination," mimeograph, Yale University 2025.

Online Appendix for

What Does "Ideas Are Getting Harder to Find" Mean, Exactly?*

Teresa C. Fort† Nathan Goldschlag‡ Jack Liang§ Peter K. Schott¶ Nick Zolas‖

A Matching Patents to Census Business Data

This appendix section describes how we match patent assignees from the United States Patent and Trademark Office (USPTO) data to Census Business microdata files from 1977 to 2021, the latest available year of microdata.

A.1 Data Preparation and Cleaning

The BDSPF-Long match process first prepares the USPTO patent data and the combined County Business Patterns and Business Register (CBPBR) microdata files. On the patent-assignee side, we combine extracts from Google patent XML files with additional information on assignees (such as county) from the PatentsView data. We augment patent-assignee geography with location information on inventors, attaching each unique geography among a patent's inventors to the assignee. This addition allows us to search within the inventors' geography for a business name match for the assignee, in addition to the assignee's geography listed on the patent.

To reduce computational burden, we collect only unique assignee name and geography information from the patent assignee data and maintain crosswalks to facilitate matching those unique name and geography combinations (labeled by `patnameid`) back to patent-assignee records (`patnum,assg_seq`).

On the CBPBR side, we collect establishment business name and geography information (physical and mailing) from the CBPBR files, using retimed files in years with retimed data.[1] We use both the `name1` and `name2` fields from the CBPBR, along with the concatenation of the two. We match by establishment identifiers to the Longitudinal Business Database (LBD) to attach `lbdfid` and `firmid` to the CBPBR records. Since our goal is a firm-level match, we de-duplicate to keep unique firm identifier, name, and geography combinations.

*Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7083300: CBDRB-FY25-CES002-001)

†Tuck School at Dartmouth, US Census Bureau, NBER, and CEPR.

‡US Census Bureau.

§Temple University.

¶Yale School of Management, CEPR, and NBER.

‖Department of State

[1]No retimed CBPBR files exist in Economic Census years, in 1976, or years near the end of the time series if the last year of data is not an Economic Census Year. Retimed CBPBR files have been modified to add and remove establishment records based upon imputed inter-censal births and deaths of establishments associated with small multi-unit firms that are not surveyed by the Company Organization Survey (COS).

Both the CBPBR and patent assignee data are subjected to identical cleaning algorithms. We apply SAS Data Quality (DQ) name and geography standardization macros, along with a customized suite of "find and replace" string commands that remove or shorten words or phrases within the name field.[2] After standardizing and cleaning the name and geography information, we generate DQMATCH codes to facilitate fuzzy matching.[3] We also generate partial name fields, one containing the first eight characters, and another containing the first two words of the business name, along with a "stop word" name field. The "stop word" name removes common words and phrases from the name field.[4] Finally, we create a version of the name that removes all spaces from the string.

To summarize, after the data preparation steps we have several versions of the CBPBR and assignee name fields, namely:

1. Raw name (`name`)
2. Standardized name (`name_std`)
3. Stop word name (`name_stp`)
4. No space name (`nameNS`)
5. DQ90 match code name (`name90`)
6. First two words name (`name2Word`)
7. First 8 characters name (`name8Char`)

We also have the following geography variables:

1. State (`state`)
2. 3-character zip code (`zip3`)
3. 3-character FIPS county (`cty`)
4. Standardized city (`city`)
5. DQ90 match code city (`city90`)

A.2 Name and Geography Matching

Each unique name and geography combination collected from the patent data is matched to every unique Census `firmid`, name, and geography combination in every year covered by the CBPBR. In the post-match cleaning steps described below, we implement an allowable gap between the CBPBR year and grant year of a given patent. We use the patent's grant year since the USPTO assignee name is based on the firm at the time of grant.

[2]For example, we change several variations of "COOPERATIVES", "COOPERATIVE", and "CO OPERATIVE" to simply "COOP".

[3]DQMATCH codes use proprietary algorithms to generate hashcodes that group together similar strings. Two records will be given the same hashcode if their string characteristics are sufficiently similar. The SAS DQMATCH function allows for varying levels of "fuzz" in how distant strings are allowed to be before being assigned different hashcodes.

[4]For example, in the "stop word" name we might remove the word "INDUSTRIES" from the name field.

Our name and geography matching is organized into blocks and passes. A match pass identifies records in the CBPBR and patent assignee data that agree on a given set of criteria. A block consists of one or more match passes with broadly similar match criteria. We residualize after each block, removing patent-assignee records that received at least one match, before proceeding to the next match block. Match passes are incremented and assigned to matches as they are made. Organizing our match passes in this way allows us to: (1) distinguish between different tiers of match quality across match blocks, and (2) account for ambiguity in which match pass is of higher quality within match blocks.

The match blocks and passes are organized as follows:

- Block 1: three passes by full name and 2 geographic elements
    1. Name, city, state
        (a) (name), `city_std`, `state`, where (name) includes [`name`, `name_std`, `name_stp`, `nameNS`, `name90`]
        (b) `name90`, `city90`, `state`
    2. Name, zip code, state
        (a) (name), `zip3`, `state`
    3. Name, county, state
        (a) (name), `cty`, `state`
- Residualize, removing patent assignee name and geographies matched in block 1
- Block 2: three passes by full name and 1 geographic element
    1. Name, City
        (a) (name), `city_std`, where (name) includes [`name`, `name_std`, `name_stp`, `nameNS`, `name90`]
        (b) `name90`, `city90`
    2. Name, zip code
        (a) (name), `zip3`
    3. Name, state
        (a) (name),`state`
- Residualize, removing patent assignee name and geographies matched in block 2
- Block 3: 1 pass by full name but no geographic elements
    1. Name
        (a) (name), where (name) includes [`name`, `name_std`, `name_stp`, `nameNS`] (note `name90` is excluded)
- Residualize, removing patent assignee name and geographies matched in block 3

- Block 4: 1 pass by incomplete name and no geographic elements

    1. Partial name

        (a) (name), where (name) includes [`name2Word`, `name8Char`]

In Block 1, match pass 1, for example, we identify CBPBR and patent-assignee records that match on raw name (`name`), standardized city (`city_std`), and state (`state`). When a match is stored, we also compute the string lengths and a series of string comparators between the two name fields. These string comparators include: Levenstein distance (`lev`), Jaro-Winkler (`jw`), SAS compare (`cmp`), and generalized edit distance (`ged`). We use these string comparator measures, along with the match pass, in subsequent post-match cleaning algorithms to disambiguate matches.

A.3 Match Cleaning and Longitudinal Imputation

After the patent assignee name and geography information has passed though all match blocks for all years of the CBPBR, we perform a series of post-match cleaning steps that balance several competing goals. We aim to (1) eliminate ambiguity in the match, leaving each patent-assignee record (`patnum`, `assg_seq`) matched to a single firm (`firmid`), (2) limit the presence of incorrect matches (limit false positives), and (3) match as many patent assignee records as possible (limit false negatives). As with many matching exercises, there is an inherent tension between these goals. Additional matches can always be made by lowering quality thresholds, trading off fewer false negatives for more false positives.

A novel attribute of our match strategy is that we leverage disambiguated assignee identifiers available in the USPTO's PatentsView data (`pv_asgid`).[5] PatentsView is a patent data visualization and analysis platform that transforms patent documents stored in XML into a relational database with disambiguated identifiers. In the source data, the USPTO does not track unique inventors or assignees across patent documents. The same patent assignee may appear on multiple patents with slight differences in its reported name or geography. The PatentsView data implements clustering algorithms that group patent assignee records (`patnum`, `assg_seq`) that have similar name and geography information. We utilize those identifiers to rationalize `firmid` matches across patents. For instance, if the PatentsView data groups together multiple patents as belonging to the same assignee, and our matching algorithm links only a subset of those patents, we leverage the PatentsView `pv_asgid` to impute the `firmid` from matched to unmatched patents.

Before analyzing the sequence of `firmid` matches for a given assignee, we clean and disambiguate the raw matches. Details of these cleaning steps are described in Section A.5. The goal of the match cleaning steps is to create an unbalanced patent assignee-year (`pv_asgid`, `grant_yr`) panel with a single matched `firmid` for each observation. We summarize the different cleaning steps here.

1. Remove matches outside of +/- 2 years from `grant_yr`.
2. Remove matches to firms not payroll active in `grant_yr` or `grant_yr`-1.

[5] In the version of PatentsView used here, the disambiguated assignee identifier is labeled assignee_id. We label this pv_asgid throughout.

3. Select among multiple matches based upon string comparators.
4. Select most frequently matched `firmid` to an assignee (`pv_asgid`) within a given `patnum`, `assg_seq`.
5. Select most frequently matched `firmid` within a pooled set of years (+/- 2 years).
6. Select most frequently matched `firmid` within a given `pv_asgid`, `grant_yr`.

If after applying these match cleaning steps there is more than one match for a given patent assignee-year (`pv_asgid`, `grant_yr`), we drop all matches for that patent assignee-year. We do this because the longitudinal algorithms that operate at the patent-assignee-year level require unique matches. Note that each patent assignee-year maps to one or more patents granted to that assignee in that year. In cases in which we remove all matches due to an inability to resolve to a single matched `firmid`, we rely on the longitudinal imputation algorithms to identify a match. At this point, prior to the longitudinal imputation processing, we have a panel of all unique [`pv_asgid`, `grant_yr`] combinations with zero or one matched `firmid`.

The goal of the longitudinal imputation is to use the sequence of matches for a given assignee (`pv_asgid`) over time to identify additional matches. For instance, if a patent assignee is missing a match in its first year in the panel, but has a match in its second year, we can push the `firmid` matched in the second year "backwards" in time to the first year. There are several types of longitudinal imputations that copy matches "forward" and "backward" in time and fill gaps between matches. To make decisions about which matches can be imputed, we need to identify, for each patent assignee, the first and last non-missing match and for each patent assignee-year, the closest non-missing match, both forward and backwards in time. After filling gaps based on leading and lagging matches, we assign missing matches to the modal match across years for a given assignee. The longitudinal imputations are then performed as follows:

1. *Leading gaps* For assignees with missing assignee-year records *prior* to their first matched patent, set the `firmid` equal to the `firmid` from their first matched year.
2. *Lagging gaps* For assignees with missing assignee-year records *after* their last matched patent, set the `firmid` equal to the `firmid` from their last matched year.
3. *Interior gaps* For assignees with missing patent-year records in between matched years, set the missing `firmid`s equal to the closest lagged `firmid` when the closest lag and lead `firmid` are identical (e.g. AA...A).
4. *Complex interior gaps* When the closest lag and lead `firmid`s are not the same:
   (a) Set `firmid` to the closest lag `firmid` if the year is closer to the lag than the lead.
   (b) Set `firmid` to the closest lead `firmid` if the year is closer to the lead than the lag.
   (c) Set `firmid` to the closest lag `firmid` if the year is equally close to the lead and lag.
5. *Global modal* Set any unmatched [`pv_asgid`, `grant_yr`] combinations to the modal match `firmid` for the `pv_asgid` across years.

After the longitudinal processing is complete, we move from the assignee-year level (`pv_asgid`, `grant_yr`) back to patent-assignee records (`patnum`, `assg_seq`). We verify that all `firmid` matches have positive payroll in $t-1$ and/or t (payroll denom positive) in the LBD within +/- 2 years of the `grant_yr`. Since a matched `firmid` need not be in the LBD in the `grant_yr` (e.g. not payroll denom positive in the `grant_yr`), we include `lbd_yr` in the crosswalk, which captures the year of LBD to which that the patent-assignee record can be matched.

A.4 Analysis of BDSPF-Long Crosswalk

In this section, we present a series of analyses that demonstrate the properties of the BDSPF-Long crosswalk. These analyses includes how it compares to other patent assignee-firm crosswalks and how alternative cleaning algorithms affect the quality of matches.

Match Quality

We first provide an overview of match rates. Table A.1 shows the percent of US and foreign patent-assignee records (patent number-assignee sequence pairs) by match type between 1976 and 2021.[6] First, about 92% of the 3.3 million US patent-assignee records receive a raw match in the name and geo matching described in Section A.2. This share falls to 43% for foreign assignees. Note that it is not clear what percent of patent-assignees *should* match to the CBPBR, especially for foreign patents, which account for about half of patent-assignee records by the end of our data. Foreign assignees without an employer establishments in the US, for example, will not match to Census microdata. About 74% of US patent-assignee records and 21% of foreign assignee records receive a unique raw match. These statistics indicate that foreign patent-assignee matches tend to be lower quality and require additional disambiguation. Finally, our crosswalk matches 92% of US patent-assignee records and 58% of foreign assignees. Combined, these rates yield a match rate of approximately 75% for all patent-assignee records from 1976 to 2021.

Table A.1: Patent-Assignee Match Rates

% of Patent Assignees	US	Foreign	All
Raw Match	92.43	42.94	67.96
Unique Raw Match	74.39	20.93	47.96
Match in Crosswalk	92.23	57.75	75.18
Total	3,336,000	3,263,000	6,599,000

Source: BDSPF-Long, PatentsView

Note: Table shows the percent of US (has US location) and Foreign (no US location) patent-assignee records from 1977 to 2021 with at least one raw match ("Raw Match"), with a unique raw match ("Unique Raw Match"), and with a match in the final crosswalk ("Match in Crosswalk").

To provide additional detail on the composition of the raw matches, Table A.2 shows the count of

[6] US patent-assignee records are those that are associated with a city state combination in the United States.

raw matches by block. Blocks 1 and 2 condition on at least some form of geographic matching (e.g. city, state, 3-digit zip code), block 3 is business-name only matching, and block 4 is partial name matches. The second and third columns show the count and percent of raw matches made in each block. The third and fourth columns show the count and percent of patent-assignee records. The last two columns show the share of raw matches that are unique, and the percent of patent-assignee records that receive a raw match from a given block that also receive a final match in our crosswalk, or what we call the "conversion rate", of raw matches in a given block.

Overall, our 6.6 million matched patent-assignee records are based on 15 million raw matches. Roughly 31% of raw matches come from block 1, our highest quality set of match passes. Most of our matches (55%) are made in block 3 where we match on name only. Despite only 31% of raw matches made in block 1, over 41% of our final patent-assignee matches arise from block 1. This high share arises because the raw matches are higher quality, with fewer false positives, and thus more likely to be unique. Indeed, the unique rate for block 1 matches is almost 84%.

About 3% of our matched patents come from the block 4, partial name match. The unique rate declines significantly across blocks as the match criteria become weaker, consistent with higher blocks having more false positives. The conversion rate remains high for all blocks. Over 98% of patent assignee records that receive a block 1 raw match end up with a match in the crosswalk. This high rate is consistent with the fact that block 1 matches tend to be higher quality. The lowest quality block, block 4, still has a conversion rate of nearly 88%. The final row of Table A.2 highlights a key contribution of our new crosswalk. 35% of patent-assignee records that did not receive a raw match nevertheless have a final match in the crosswalk. These final matches are due to the cleaning steps that aggregate across patents within an assignee (`pv_asgid`), and to our longitudinal imputations.

Table A.2: Raw Match Block Distribution

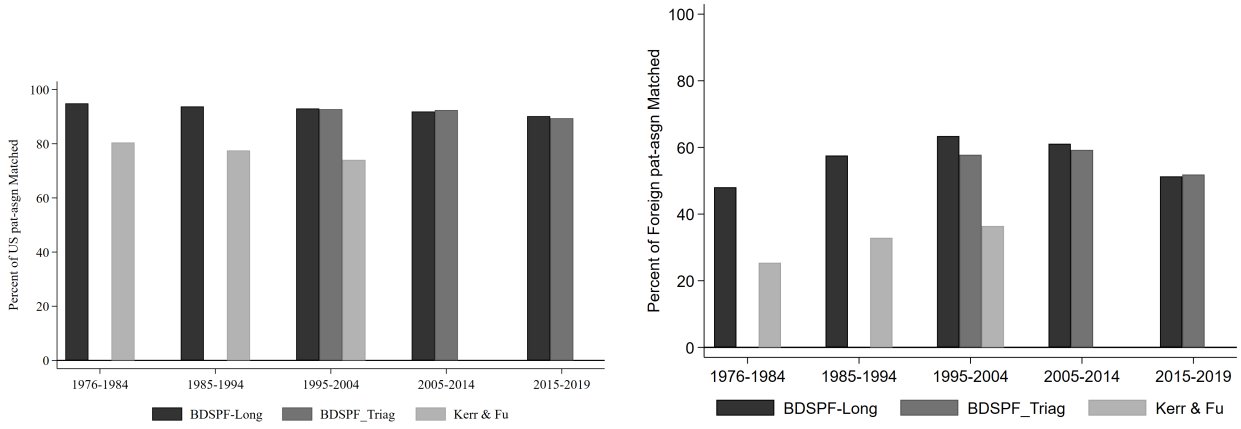
	Raw Match		Final Match		Unique	Conversion
Block	Count	Percent	Count	Percent	Rate	Rate
1	4,786,000	31.04	2,728,000	41.33	83.87	98.24
2	364,000	2.36	182,000	2.76	75.27	95.6
3	8,551,000	55.45	1,386,000	21	48.41	87.09
4	1,720,000	11.15	190,000	2.88	36.05	87.89
None	0	0	2,114,000	32.03		34.67
Total	15,421,000	100	6,600,000	100	47.95	75.17

Source: BDSPF-Long, PatentsView

Note: Table shows the count and distribution of raw matches ("Raw Match") by match block. A single patent-assignee record may have multiple raw matches. The "Final Match" columns present the count and percent of patents by the type of raw match used in crosswalk. "Unique Rate" shows the percent of patents with a given raw match type that had a unique raw match. "Conversion Rate" shows the percent of patents with a raw match that are also in the final crosswalk (e.g. the raw match was "converted" into a final match). Block 1 matches use name, city, and state. Block 2 matches use name and city. Block 3 matches use name only. Block 4 matches use an incomplete name match. The positive conversion rate and final matches for patent-assignee records with no raw match is due to assignee matches and longitudinal imputations.

Figure A.1 shows the match-rate time series of the BDSPF-Long crosswalk alongside the Kerr and Fu (2008) (Kerr & Fu) and the BDSPF-Triangulation crosswalk (Graham et al., 2018; Dreisigmeyer et al., 2018). The time coverage varies significantly between the two other crosswalks. Kerr & Fu covers 1976 to 2001, while the BDSPF-Triangulation, due to the availability of LEHD data, covers from 2000 forward. As noted previously, one advantage of BDSPF-Long is the long time horizon covered by the matches, 1976 to the most recent LBD year. The BDSPF-Triangulation crosswalk only goes back to 2000 due to coverage limitations of the LEHD employee-employer matched data. The US assignee-record match rate of BDSPF-Long is similar to the BDSPF-Triangulation crosswalk rate and significantly higher than Kerr & Fu rate. Both BDSPF-Long and Kerr & Fu exhibit increasing match rates for foreign-assignee records between the late 1980s and the early 2000s, and both BDSPF-Long and BDSPF-Triangulation match rates decline from the mid 2000s for foreign-assignee records.

Figure A.1: Match Rate Time Series



(a) US Patent Assignees

(b) Foreign Patent Assignee

Source: BDSPF-Long, PatentsView

Notes: Panel (a) shows the percent of patent-assignee records, granted within a given time window, that received a match in each crosswalk. BDSPF-Long is the crosswalk introduced in this paper. BDSPF-Triag is the LEHD-based crosswalk described in Graham et al. (2018) and Dreisigmeyer et al. (2018). Kerr & Fu is the crosswalk developed by Kerr and Fu (2008).

Table A.3 shows the count of `firmid`s per `pv_asgid` in the raw match. There are several reasons the number of `firmid`s per `pv_asgid` may be greater than one. For example, mergers and acquisitions may change the firm identifier associated with a patent assignee over time in the CBPBR. About 62% of patent assignees (`pv_asgid`) do not receive a raw match at all. Given that 67% of patent-assignee records get a raw match, shown in Table A.1, this suggests that the assignees that do not get a matched `firmid` account for relatively few patents. About 25% of assignees are matched to a single `firmid`. Moreover, the cleaning steps reduce the percent of `pv_asgid`s with 10 or more `firmid`s dramatically, from 2.4% to nearly zero. Some of this reduction comes from simply removing matches that are too ambiguous. The percent of `pv_asgid`s with no `firmid` matches rises from 62% in the raw match to 65% in the final crosswalk. These results suggest that the cleaning steps significantly reduce the ambiguity of matches within `pv_asgid`s.

Table A.3: Match Ambiguity

	Raw Match		Final Match	
Firmids per Assginee	Count	Percent	Count	Percent
0	353,000	61.81	374,000	65.41
1	145,000	25.39	169,000	29.55
2	30,500	5.34	21,500	3.76
3	11,500	2.01	5,000	.87
4	6,100	1.07	1,400	.24
5-9	11,500	2.01	900	.16
10+	13,500	2.36	20	0

Source: BDSPF-Long, PatentsView

Note: Table shows the count and distribution of patent assignees (`pv_asgid`) that have a given number of matched `firmid`s across all of their patents in the raw match ("Raw Match") and in the final crosswalk ("Final Match").

The cleaning steps also increase the consistency of matches within a `pv_asgid` across years. Table A.4 reports the percent of `pv_asgid`s and `pv_asgid`-years for which there is a change, from year-to-year, in the `firmid` matched. Note that, by construction, the final crosswalk matches a single `firmid` to a given `pv_asgid`-year, but the raw match may have multiple `firmid`s per `pv_asgid`-year. As such, in the raw match we consider it a "break" if the set of `firmid`s is different between $t-1$ and t. In the raw match, 8.2% of `pv_asgid`s had at least one break, or change in `firmid`, across the time series. In the final crosswalk this falls to 5%. In terms of `pv_asgid`-years, 7.9% exhibit year-to-year breaks in their matched `firmid`s in the raw match compared to 3% in the final crosswalk. Again, we may not expect the count of breaks in matched `firmid`s to be zero due to mergers and acquisitions. Despite this expectation, our cleaning algorithms significantly reduce the probability that a `firmid` changes from year-to-year for a given `pv_asgid`.

Table A.4: Breaks in PV Assignee ID-Firmid Over Time

	Raw Match	Final Match
Assignees	571,000	571,000
Assignees with Breaks	47,000	28,500
Assignees with Breaks %	8.23	4.99
Assignee-Years	1,478,000	1,478,000
Assignee-Year Breaks	117,000	45,000
Assignee-Year Breaks %	7.92	3.04

Source: BDSPF-Long, PatentsView

Notes: Table shows the count of patent assignees and patent-assignee year combinations that experience a change in their matched `firmid` between years in the raw match ("Raw Match") and in the final crosswalk ("Final Match").

While the cleaning steps reduce ambiguity among raw matches, the longitudinal imputations increase match rates by extending the reach of the clean name and geography matches we make. Table A.5 shows the count and percent of matched patent assignee records by the longitudinal imputation status. About 7% of matches are identified using longitudinal imputes, which account for about 5.8% of all records. The majority of longitudinal-impute cases are "Lagging Gaps", where we push matches forward in time to unmatched `pv_asgid`-years. The next most frequent types of longitudinal imputations are Leading Gaps (pushing matches backwards in time) and complex interior matches (filling gaps with different matches on either end).

Table A.5: Longitudinal Imputations

Longitudinal Imputation	Count	Percent
Unmatched	1,396,000	21.16
None	4,827,000	73.16
Leading Gap	110,000	1.67
Lagging Gap	133,000	2.02
Simple Interior	49,500	.75
Complex Interior	82,500	1.25
Total	6,598,000	100

Source: BDSPF-Long, PatentsView

Note: Table shows the count of patent-assignee records (`patnum`, `assg_seq`) by the type of longitudinal impute they received, if any.

To assess the quality of our matches we compare them to the very high quality BDSPF-Triangulation model "A1" matches. These matches are ones for which the BDSPF-Triangulation algorithms are able to "fully traingulate" an inventor and assignee, both being corroborated by employment records in LEHD. There are approximately 1.4 million "A1" matches. Among the BDSPF-Triangulation "A1" matches, we compute the number of "correct" BDSPF-Long matches (those for which both crosswalks agree), the number of incorrect BDSPF-Long matches (those for which the crosswalks disagree) and the number of records for which BDSPF-Long is unable to identify a match.

We measure precision of the BDSPF-Long matches as the number of correct matches divided by the sum of correct and incorrect matches. This metric captures the share of matches we make that are correct. We then compute recall as the count of correct and incorrect matches divided by the count of correct, incorrect, and missing matches. This metric captures the share of records for which we make a match, correct or not.

Table A.6 reports the results of this analysis. The precision of matches for US assignees records is significantly higher than for foreign assignee records (94% vs. 81%). Recall shows an even bigger gap between US and foreign assignees (98% vs 80%). The BDSPF-Triangulation model "A1" matches have certain properties that make them unrepresentative of the typical patent assignee record (e.g. they are more likely to be US-based). However, at least among the US-based patent assignee records,

our matches tend to be very high quality within this subset.

Table A.6: BDSPF-Long Match Precision & Recall

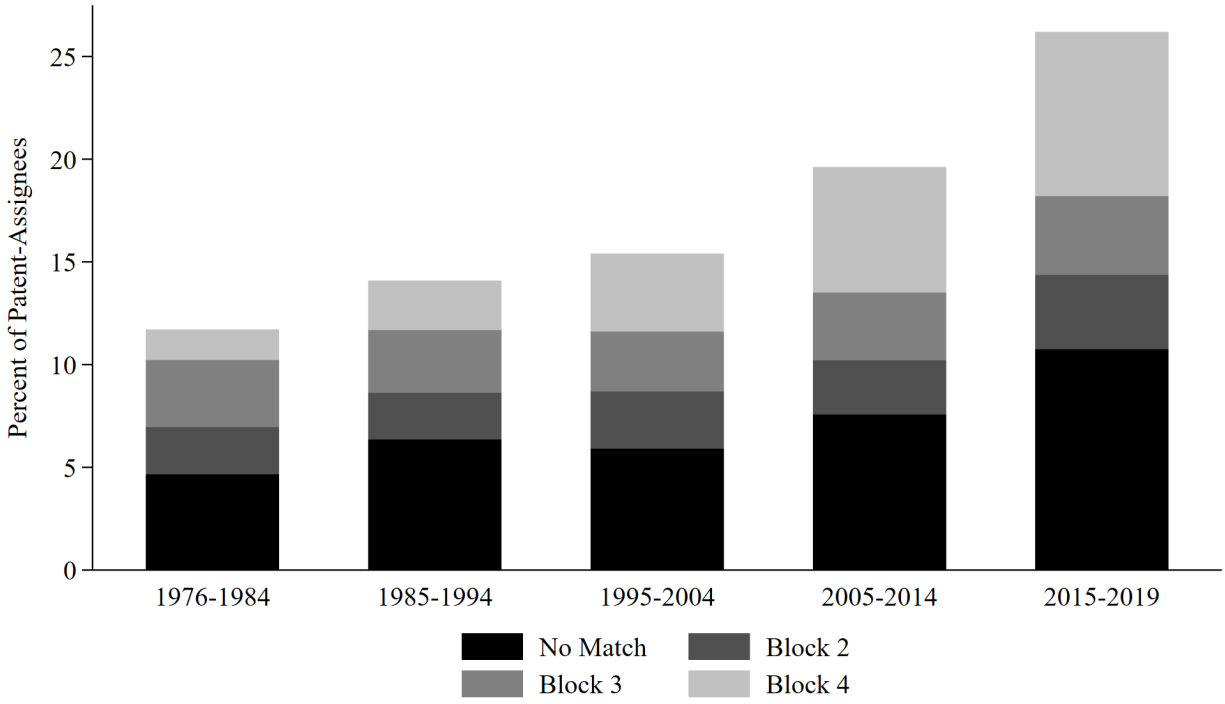
	US	Foreign	All
Precision	93.6	80.98	93.08
Recall	97.94	80.44	97.23
Total pat-asgn Records	1,350,000	57,500	1,407,500

Source: BDSPF-Long, PatentsView

Notes: Table shows the precision and recall of BDSPF-Long matches using the "A1" model (e.g. highest quality) matches from the BDSPF-Triangulation crosswalk as a truth set. Precision is the count of patent-assignee records (`patnum`, `assg_seq`) with agreement between BDSPF-Long and BDSPF-Triag divided by the count of all records with for which the BDSPF-Long has a matched `firmid`, regardless of agreement. Recall is the count of patent-assignee records with agreement between BDSPF-Long and BDSPF-Triag divided by the sum of the count of matched records with agreement and the records with no BDSPF-Long match.

Finally, we turn to the declining match rate for US-based assignees after 2010. The match rate for US assignees falls from 95% to 90% between the early 1980s and the late 2010s. As noted above, it is difficult to interpret patent-assignee match rates because the true subset of employer businesses among patent assignees is unknown. Despite this limitation, we can evaluate whether there are changes over time in the prevalence of certain types of matches. Figure A.2 shows the decline the percent of patent assignees by raw-match type over time. The excluded category, which when summed with the displayed categories sums to 100%, is block 1 matches. Several patterns in Figure A.2 are worth noting. First, the share of US patent-assignee records with no raw match to the BR rises from 5% to 11%. Second, there has also been an increase in the share of block 4 matches, which are the lowest quality partial name matches. Third, we see a substantial decline in the excluded block 1 category, which falls from 88% to 74%.

Figure A.2: US Patent Assignee Records by Raw Match Status



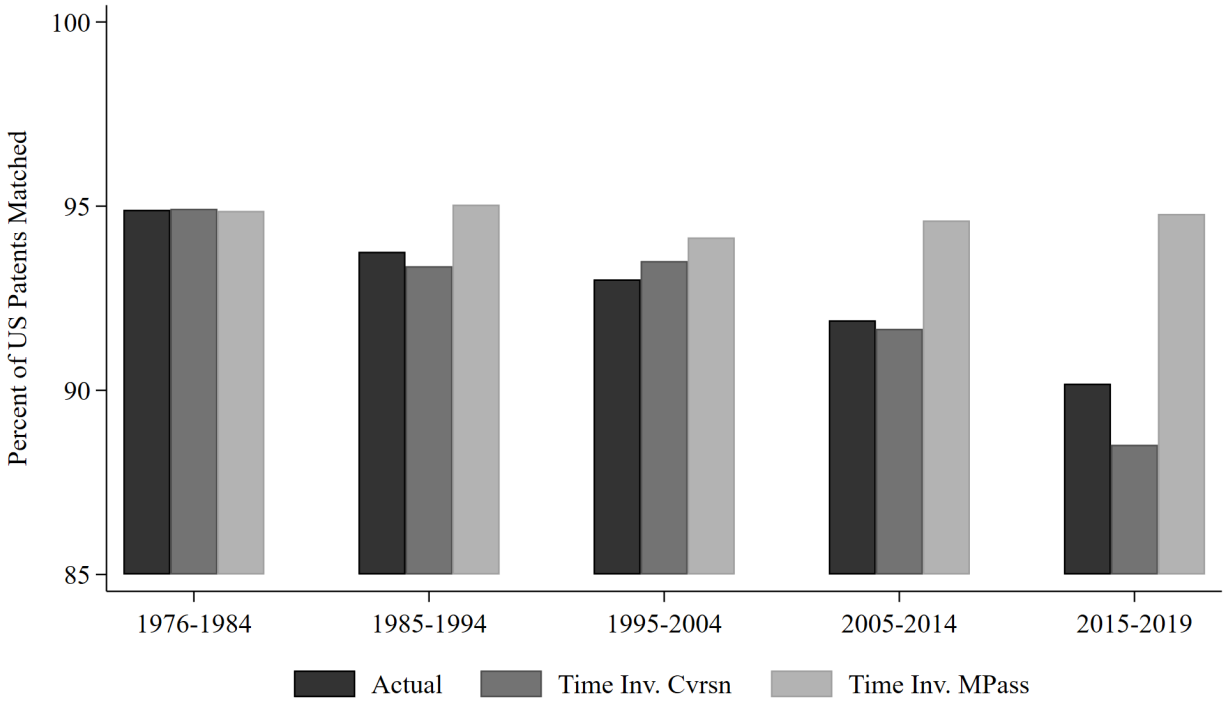
Source: BDSPF-Long, PatentsView
Notes: Figure shows the distribution of patent-assignee records (`patnum`, `assg_seq`) by the type of raw match, if any. The excluded category, which is also the largest, are those with a Block 1 match. If the Block 1 percent were included all groups would sum to 100 within a given time period.

The changing composition of raw-match status among patent-assignee records may have a significant effect on match rates if our ability to identify a final matched `firmid` for a given record varies with the type of raw match made, which appears to be the case in last column of Table A.2. Indeed, approximately 98% of `pat-asgn` that receive a Block 1 match have a final match compared to 88% for Block 4.

Knowing that match quality varies by raw-match block, and that the composition of raw matches changes over time, suggests that the changing raw match composition may explain the declining match rate for US-based assignees. Figure A.3 assesses this possibility by estimating a counterfactual match rate in which either "conversion rates" are held constant at initial levels, or match-pass composition is held constant. The first set of bars shows the actual US-based patent assignee record match rate, as depicted in the left panel of Figure A.1. The second bar in Figure A.3 shows what match rates would have been if the conversion rates by match pass were held at their initial levels and match-pass shares evolved as they do in the underlying data. The third bar in A.3 shows what match rates would have been if the match-pass composition was held at its initial level and conversion rates varied by year as they do in the underlying data.

As we can see, the decline in match rates would be exacerbated if conversion rates were held constant at their initial levels. In contrast, if the match-pass composition had remained constant, match rates for US-based assignees would have stayed unchanged at about 94%. This exercise suggests that the composition of raw matches explains the decline in match rates. Over time, more raw matches are made in block 4 or not at all, putting downward pressure on match rates.

Figure A.3: Counterfactual Match Rates



Source: BDSPF-Long, PatentsView
Notes: Figure shows actual and simulated match rates for the BDSPF-Long crosswalk over time. The "Time Inv. Cvcrsn", or time-invariant conversion rate measure, shows what match rates would be if raw match composition across match passes vary as they do in the data but conversion rates (the rate at which matches from each block are converted into final matches) remains as they were in the 1976-1984 period. The "Time Inv. Mpass", or time invariant match pass composition measure, shows what match rates would be if conversion rates for each mar match pass vary as they do in the data, but the raw match composition remains as they were in the 1976-1984 period.

Match Experiments

As is often the case when matching data to the Business Register, the matching procedures are complex and reflect choices made at various stages of development. To guide these choices, we conducted several "experiments" in which a single component of the matching procedure was changed. We then compared the precision and recall among the BDSPF Triangulation A1 cases, as in Table A.6, for each experimental crosswalk.

The first experiment used application year to window matches initially, instead of relying only on the grant year. If an assignee were active around the application of the patent, but for one reason or

another exits before the grant year, perhaps due to an acquisition, it might be missed using only the grant year. In this case, since app-to-grant gaps average roughly 3 years, windowing only around the grant year is too restrictive and excludes true positives.

The second experiment focused on a more fundamental question about the unit of processing during the longitudinal processing steps. Rather than aggregate across patents to the `pv_asgid`-year level, one could instead perform all processing steps at the patent-record level. This approach naturally allows for some within `pv_asgid`-year variation in firmid matches, since it does not impose a single `firmid` match for the `pv_asgid`-year, which are then pushed back down to the patent-assignee records.

Table A.7 presents the precision and recall statistics for these two experiments. We focus only on US-based assignees. The precision for both experiments is slightly lower than the preferred algorithm reported in Table A.6 (93.6% for US-based assignees). The recall for both are slightly higher than the preferred algorithm (97.9%). Even with the more fundamental change of unit of processing in the second experiment, there is very little change to the overall quality of the match.[7]

Table A.7: Experiments Precision & Recall

	Preferred Algorithm	Experiment 1 Application Year	Experiment 2 Patent-Level Processing
Precision	93.6	90.97	91.03
Recall	97.94	99.53	99.51

Source: BDSPF-Long, PatentsView

Table shows the precision and recall of US-based BDSPF-Long matches using the "A1" model (e.g. highest quality) matches from the BDSPF-Triangulation crosswalk as a truth set. The first column ("Preferred Algorithm") replicates the US-based assignee record precision and recall measures from Table A.6. The second and third show similar statistics from two different experiments where the preferred algorithm was modified by either using application year instead of grant year ("Experiment 1 Application Year") or processing was done at the patent-level instead of the assignee-year-level ("Experiment 2 Patent-Level Processing"). Precision is calculated as count of patent-assignee records (`patnum`, `assg_seq`) with agreement between BDSPF-Long and BDSPF-Triag divided by the sum of the count of records with agreement and those that disagree but for which the BDSPF-Long has a matched `firmid`. Recall is calculated as the count of patent-assignee records with agreement between BDSPF-Long and BDSPF-Triag divided by the sum of the count of records with agreement and those that did not have a BDSPF-Long match.

A.5 Match Cleaning and Longitudinal Imputation Detail

The match cleaning and longitudinal imputation program proceeds according to the following steps.

1. Collect PatentsView data.

[7] One experiment resulted in a change to the preferred algorithm. That experiment removed a step after 6.a. in Section A.5 that kept only matches with the lowest match pass. Removing this constraint, i.e., retaining all match passes within a block, increased precision considerably, so we retain all passes in the final algorithm. This match-quality increase suggests that the relative quality of matches varies across match blocks, but that within a block, the match pass is not a strong predictor of quality.

(a) **pv_ayr**: [patnum, app_yr]

(b) **pv_gyr**: [patnum, grant_yr]

(c) **pvassigneeid**: [pv_asgnid, patnum, assg_seq, app_yr, grant_yr]

(d) **pvassgineeid_frame**: [pv_asgnid, grant_yr], unique [pv_asgnid, grant_yr] combinations in **pvassgineeid**.

(e) **pvpatasgn**: [patnum, assg_seq, pv_asgnid, rawlocation_id]

(f) **pvloc**: [rawlocation_id, location_id, city, state, country, latlong]

2. Create a frame of all patents as a backbone for imputation.

   (a) Start with a list of all patent-assignee sequence combinations, **pvpatasgn**.

   (b) Add location information, merge (m:1) by rawlocation_id to **pvloc**.

   (c) Add app and grant years, merge (m:1) by patnum to **pv_ayr** and **pv_gyr**.

   (d) This yields **pvMatchFrame**: [patnum, assg_seq, app_yr, grant_yr, assignee_id, rawlocation_id, location_id, city, state, country, latlong]

3. Identify a window of valid years for each patnameid, the identifier used in the raw matches.

   (a) Start with **patnameid_patid** dataset genreated by the patent data prep programs, which maps unique name+geo combinations to patent assignee records.

   (b) Create a long dataset that maps patnameid to all years within +/- two years around the grant year of all patents the patnameid maps to.

   (c) For each [patnameid, year] pair that exists within +/- two years of a patent grant, we compute the minimum gap between that year and the closest patent grant.

   (d) This yields **patnameid_validyears**: [patnameid, year, mingap]

4. Identify a what years every firmid is payroll active.

   (a) Loop over the LBD establishment datasets, keeping all unique firmids in either the firmid or firmid_rorg fields. firmids that exist in a given year of the LBD necessarily have an associated payroll-denom positive establishment. We keep track of what field the firmid is drawn from.

   (b) This yields **lbd_active**: [firmid, year, source].

5. Stack raw match files.

   (a) Raw match files contain matches of every patent, via the raw match processing described above, to a given year of the CBPBR.

   (b) Loop over the years to create a single file.

   (c) This yields **allmatches**: [patnameid, year, geotype, fid, mpass, lev, jw, cmp, ged, l_strlen, r_strlen]

6. Perform first pass of cleaning **allmatches**.

   (a) Merge (m:1) by `year`, `patnameid` to **patnameid_validyears**, keeping only matched records. This limits [`patnameid`,`year`,`firmid`] matches that are within a valid window around the grant years.

   (b) Collapse to the [`patnameid`,`year`,`firmid`] level, collapsing across [`patnameid`,`year`,`firmid`] matches that were made via different names. We take the minimum of the string comparators and an average of the length of two name strings (BR `l_strlen` and patent `r_strlen`).

   (c) Normalize string comparator measures by indexing to the best measure within [`patnameid`,`year`]. This transforms the string comparator into the ratio of the value for a given match and the best (lowest) string comparator of that type (e.g. `lev`, `jw`, etc) for a given [`patnameid`,`year`].

   (d) Create combination string comparator score, `w`, as the equally weighted squared sum of the normalized measures.[8]

7. Move matches to patent assignee record level.

   (a) Joinby `patnameid` using **patnameid_patid**, keeping [`patnum`, `assg_seq`, `grant_yr`, `firmid`, `cbpbr_yr`, `mpass`, `w`].

   (b) Now that we are at the [`patnum`, `assg_seq`], we again limit to +/-2 years from grant.

   (c) Exclude matches ([`firmid`,`cbpbr_yr`]) that are not active within +/-2 around the grant year. To do this, we merge (m:1) by `firmid`, `grant_yr`=`year` to **lbd_active**, keeping only matched records.

   (d) Again, at the `patnum`, `assg_seq` level we keep only the best match pass for a given patent. Above we took the best match pass at the `patnameid`, `year` level, but a patent may have multiple `patnameid`s associated with it (e.g. more than one geo). Within a `patnum`, `assg_seq`, we keep only matches assocaited with the lowest match pass for the `patnum`, `assg_seq`.

   (e) Add PatentsView assignee identifiers, merge (m:1) `patnum`, `assg_seq` to **pvassigneeid** to get `pv_asgid`.

8. More cleaning before moving to assigneeid-year level.

   (a) Drop matches based in the string comaparator index. If the gap between the best and second best match for a `patnum`, `assg_seq` is ≥ 1.2 then keep only the best match.

   (b) Select modal [`firmid`, `pv_asgid`] match within `patnum`, `assg_seq`. To do this, we count the total number of times each [`firmid`, `pv_asgid`] is matched, call that `mcount`. We then find the maximum `mcount` by [`patnum`, `assg_seq`, `pv_asgid`], which is the most often matched `firmid` for the `pvasgid` among those `firmid`s match to a given [`patnum`, `assg_seq`].

[8] For string comparators x_1, x_2,...,x_n, the combined value $w = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$.

(c) Collapse to [pv_asgid, firmid, grant_yr] level, taking the count of times an pv_asgid matches each firmid within a year and the minimum (best) string comaparator score for that [pv_asgid, firmid, grant_yr] combo and counting the number of matched patents (mtch_patcount) for triplet.

(d) Complete the panel, merge (m:1) by pv_asgid, grant_yr to **pvassgineeid_frame** to collect all valid (grant active) [pv_asgid, grant_yr] pairs.

(e) This yields a panel dataset with: [pv_asgid, grant_yr, firmid, mtch_patcount, w]

9. Save a "global modal" for use later.

   (a) At this point we compute the modal firmid for each pv_asgid, save as modalfid.

   (b) This yields **global_modal**: [pv_asgid, modalfid].

10. Compute "local modals" for use later.

    (a) Use a rolling window of +/-2 around a focal year, pool all [pv_asgid, grant_yr, firmid] matches within that window, and compute the modal firmid for each pv_asgid, associating that "local modal" (local_modalfid).

    (b) This yields **local_modal**: [pv_asgid, grant_yr, local_modalfid].

11. Again use the string comparators to drop duplicates.

    (a) Above we dropped records based on the best and second best string comparator scores by [patnum, assg_seq]. Here we look within [pv_asgid, grant_yr] pairs. If the gap between the best and second best match for an assigneeid-year is ≥ 1.2 then keep only the best match.

12. Select "local modals".

    (a) Limit to local modals, merge (m:1) by pv_asgid, grant_yr to **local_modal**, and if the count of matched firmids for a given assigneeid-year is > 1, then replace it with the local modal firmid.

13. Select patent-count modal within assignee id-year

    (a) Use matched patent count for each [pv_asgid, grant_yr, firmid] combination to select a single match by [pv_asgid, grant_yr], keeping only the firmid with the most matched patents.

14. Select modal by assigneeid-firmid within assigneeid-year.

    (a) We count the number of matches (mtch_patcount) by [pv_asgid, firmid], call that pvfidmcount. Find the firmid with the maxmimum pvfidmcount within a given [pv_asgid, grant_yr].

15. Drop duplicates within assigneeid-year.

(a) At this point we've done everything we can to select a single firmid for a given [pv_asgid, grant_yr] combination. If we still have more than one firmid for an [pv_asgid, grant_yr], we remove them all. We keep one record for the any [pv_asgid, grant_yr] that loses matches, leaving the matched firmid blank to allow longitudinal edits to fill in the match if possible.

16. Find the first and last non-missing firmid for a given assigneeid.

    (a) This yields **first_nonmissing**: [pv_asgid, grant_yr, fnmfid, fnmfid_yr], where fnmfid_yr is the year of the first non-missing firmid for a given pv_asgid.

    (b) This also yields **lst_nonmissing**: [pv_asgid, grant_yr, lnmfid, lnmfid_yr].

17. Flag longitudinal pattern types.

    (a) Flag pv_asgid as type=1 if the pv_asgid is matched to only one firmid across grant_yrs (e.g., a firmid match sequence of AAA).

    (b) Flag pv_asgid as type=3 if the pv_asgid switches back and forth to the same firmid (e.g., a match sequence of AABA).

    (c) Flag assigneeid as type=2 if the pv_asgid is not type 1 or 3, and matches to more than zero firmids (cannot be be 1, if it was then it would be type=1) (e.g., a match sequence of AAABBB).

18. Perform longitudinal imputes.

    (a) Merge on the most recent non-missing lead and lag firmids (1:1) by pv_asgid, grant_yr to **first_nonmissing** and **lst_nonmissing**.

    (b) Leading gaps: if not missing fnmfid, and missing firmid, then longimpute=1 and fill firmid with fnmfid. This pushes the first non-missing firmid backwards in time.

    (c) Lagging gaps: if not missing lnmfid, and missing firmid, then longimpute=2 and fill firmid with lnmfid.

    (d) Identify most recent lag and lead matched firmid (mrleadfid, mrlagfid) and the associated gap between a given year and that most recent match (mrleadfid_yr, mrlagfid_yr) if missing firmid.

    (e) Simple interior: if not missing mrlagfid and missing firmid, and mrlagfid and mrleadfid are the same, the longimpute=3 and fill with mrlagfid.

    (f) Complex interior (lag): if not missing mrlagfid, and missing firmid, and mrlagfid is not equal to mrleadfid, and lag gap (grant_yr-mrlagfid_yr) is less than lead gap (grant_yr-mrleadfid_yr), then longimpute=4 and fill with mrlagfid.

    (g) Complex interior (lead): if not missing mrleadfid, and missing firmid, and mrlagfid is not equal to mrleadfid, and lead gap is less than lag gap, then longimpute=5 and fill with mrleadfid.

(h) Complex interior (tie): if not missing `mrlagfid`, and missing `firmid`, and `mrlagfid` is not equal to `mrleadfid`, and lead gap is equal to lag gap, then `longimpute`=6 and fill with `mrlagfid`.

19. Fill remaining unmatched assigneeid-years with global modal by assigneeid.

    (a) We merge to **global_modal** (m:1) by `pv_asgid` to get `modalfid` and replace `firmid` with `modalfid` if `modalfid` is non-missing and `firmid` is missing.

20. Move back to patent-level, matching assigneeid-year-firmid matches onto patents by assigneeid-year.

    (a) We merge to **pvassigneeid** (1:m) by `pv_asgid`, `grant_yr` to get all [`patnum`, `assg_seq`] associated with each [`pv_asgid`, `grant_yr`]

21. Apply hand edits.

    (a) Apply both a set of hand matched string name-to-`firmid`-`grant_yr` combinations (e.g., if name is "XYZ" then `firmid` is "123") and a suite of regular expressions (e.g., if name contains "ABC" then `firmid` is "456").

    (b) Remove erroneous matches to distributed service provides. There is a list of known matching issues related to cases where firms such as cafeterias have the name of the businesses they serve in the CBPBR name fields.

22. Apply LBD active scoping conditions to the final crosswalk.

    (a) Merge (m:1) by `firmid`, `year`=`grant_yr` to **lbd_active**. If the matched [`firmid`, `grant_yr`] is found in **lbd_active** then set `lbd_yr` to `grant_yr`.

    (b) For cases that do not match (`year`=`grant_yr`), then we look for a [`firmid`,`year`] in **lbd_active** that is closest within +/- 2 years around `grant_yr`. We preference gaps, in order, [+1, -1, -2, +2].

23. The final crosswalk then contains unique [`patnum`, `assg_seq`] combinations with variables: [`patnum`, `assg_seq`, `grant_yr`, `lbdfid`, `firmid`, `lbd_yr`, `longimpute`].

B Merging Patent Data into Compustat

To match the USTPO patents to Compustat, we start with an extract of CRSP/Compustat Merged (Fundamentals Annual) downloaded from Wharton Research Data Services (WRDS) on July 9, 2024.[9]

Firms in Compustat are identified by two variables, *gvkey* and *permno*. The former is a unique number assigned to each firm, while *permno* uniquely identifiers the share class of a publicly traded

[9]We choose consolidation level C, Industry Format INDL and FS, Data Format STD, Population Source D, Currency USD, Company Status Active and Inactive for years 1950 to 2024. Though both Compustat and our patent datasets extend beyond 2016, our sample period ends in that year due to the lag between the application and patent years noted earlier.

security. Though infrequent, a *gvkey* can be associated with more than one *permno* if the firm has more than one share class.

As with the LBD, we match patents to firms in Compustat by application year, in this case by combining the patent-to-*permno* mapping in KPSS with the patent-to-*gvkey* mapping of Dyevre and Seager (2024) (hereafter DS).[10] As in our LBD match, though our matching algorithm matches by grant year, we associate patents to firms in their application year, since this timing most closely approximates the state of the firm in the year the innovation occurred. Our use of two separate mappings provides a cross check on each mapping's assignment of patents to firms, as well as their patent coverage. To make the DS mapping comparable to that of KPSS, we use the *gvkey* to *permno* concordance implicit in our downloaded Compustat file to assign one or more *permno*s to each *gvkey* in the DS mapping. (This concordance is simply all unique *gvkey-permno* pairs in the Compustat file.) We then merge the KPSS mapping into the amended DS mapping by patent and *permno*. This merge identifies patent-firm pairs that are in both mappings, as well as patent-firm pairs that appear in only one of the two mappings.[11] Our matching of patents to *permno* represents the union of these mappings.

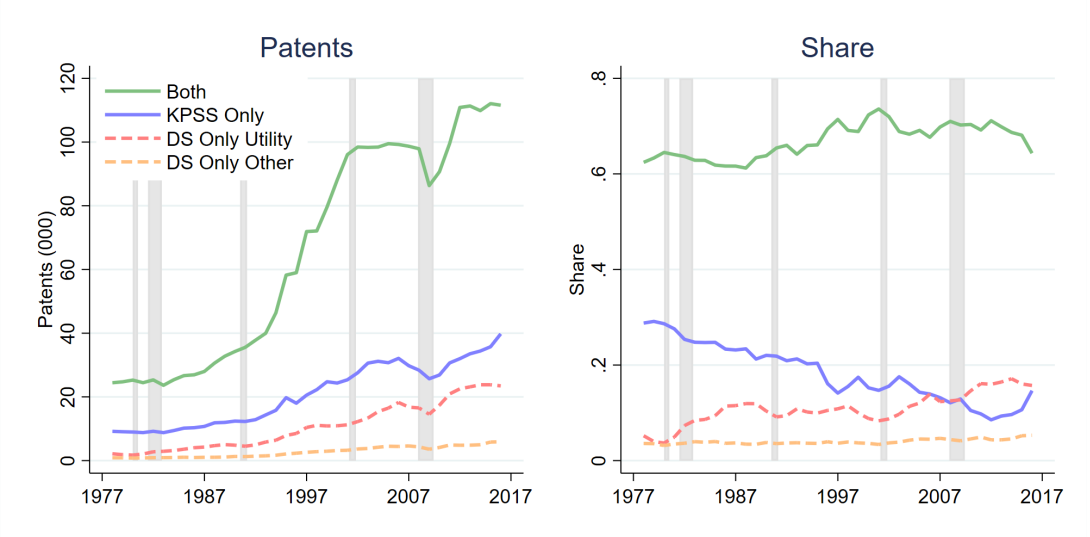
The stacked-line scatter plot in the left panel of Figure A.4 reports the cumulative number of patents matched to Compustat by either KPSS or DS in each application year, with the top line showing the total number of patents in the union of the two mappings, referred to as the "total" number of patents for the remainder of this section. As indicated in the figure, about a third of total patents appear only in one of the two datasets. One reason for this discrepancy is that DS consider all patents while KPSS (as well as KPST) focus on utility patents.[12] As indicated in the right panel of the figure, non-utility patents included only in DS (represented by the lowest dashed line) account for about 5 percent of total patents across application years. The gap between the first and second dashed lines captures utility patents contained in DS but not KPSS. These account for 5 to 15 percent of all patents across years. The distance between the highest dashed and lowest solid line picks out utility patents that are in KPSS but not DS. These account for a declining share over time, from about 30 to about 10 percent. Finally, the gap between the two solid lines captures the share of patents in both mappings, which represent about two-thirds of patents across the years.

[10] KPSS consider patents from 1926 to 2022 while DS focus on 1950 to 2020. The KPSS mapping is available on Github here. The DS concordance is available on Github here. We consider these mappings versus others in the literature because they are the most recent and because they extend to our sample period of interest.

[11] Reassuringly, across the 3,664,570 patents in the combined mappings, we find only 77,672 (about 2 percent) for which KPSS and DS differ in their assignments of a *permno*. In these cases, we default to the KPSS assignment.

[12] Utility patents account for the vast majority of patents issued by the USPTO and are for products, processes or machines that are new or improved. Other types of patents include "design" patents, which are drawings of a design with only minimal associated text, and "plant" patents for discovered or created plants.

Figure A.4: Patents in the KPSS vs DS Mappings to Compustat



Source: PV, KPSS, DS, and authors' calculations. Figure is a stacked-line scatter plot that reports number (left panel) or share (right panel) of patents matched to Compustat by KPSS and DS over our sample period, by application year. The two dashed lines represent patents utility and other patents appearing only in DS. The two solid lines capture utility patents appearing only in KPSS and in both KPSS and DS.

B.1 Match-Rate Comparisons

Table A.8 presents the match rates of patents to the LBD and Compustat panels.

Table A.8: LBD vs Compustat Match Rates

	LBD Panel		Compustat Panel	
Semi-Decade	Patents	Breakthroughs	Patents	Breakthroughs
1977	89	91	63	73
1982	88	87	60	63
1987	88	87	54	60
1992	87	88	56	67
1997	87	86	58	74
2002	86	86	60	75
2007	86	87	57	72
2012	85		55	
Average	87	87	57	70

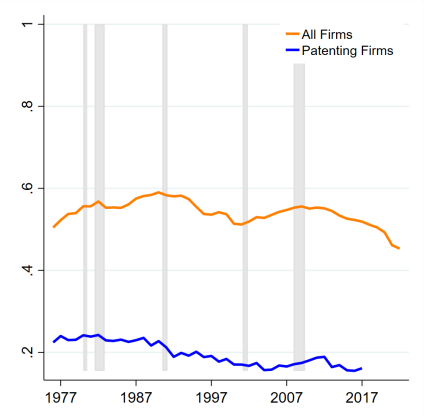
Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, RADS, Compustat and authors' calculations. Figure reports the match rates for patents and breakthroughs for the LBD and Compustat panels. These match rates correspond to the totals reported in Figure 3.

C Missing R&D

(Koh and Reeb, 2015) report substantial missing R&D cells in Compustat. In this section, we examine these missing cells in our sample.

In our raw Compustat extract, which drops firms in financial services as well as duplicate firm entries (due, for example, to firms floating more than one security) we find that 56 percent (183676/328632) of firm-year observations have missing R&D from 1950 to 2022, and 54 percent (156941/290642) have missing observations for our sample period, 1976 to 2022. If we futher restrict the sample to firms with patents from 1977 to 2022, the share of firm-year observations with missing R&D drops to 19 percent (11216/58346). Figure A.5 shows how the share of missing cells in this sample evolves over time.

Figure A.5: Distribution of Missing R&D Observations
Within Compustat Firms Over Time

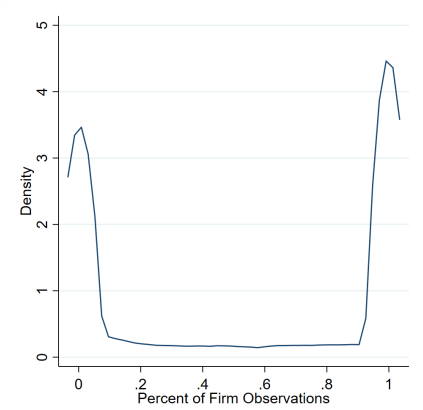


Source: PV, KPSS, DS, and authors' calculations. Figure reports the share of firm-year observations with missing R&D over 1977 to 2022 across the firms in our Compustat sample.

We compute the share of missing R&D observations within firms over time. Figure A.6 reveals that firms fall into two clusters: firms who report missing R&D for just a small share of the years in which they are present in Compustat, and those with no missing cells.[13]

[13]Examples of firms with large employment that never report R&D include Kelly Services (headhunting), McDonald's, Sears, Kroger and UPS. Examples of large employers that only partially report R&D include Toyota (missing 16/36), Unilever (.12), and AT&T (.62).

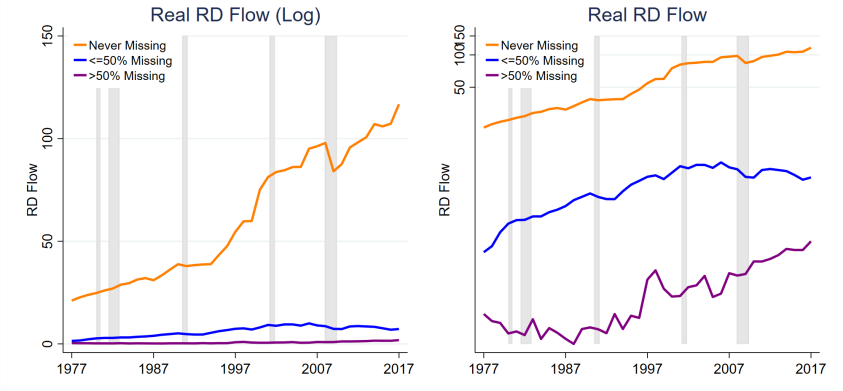
Figure A.6: Distribution of Missing R&D Observations
Within Compustat Firms Over Time



Source: PV, KPSS, DS, and authors' calculations. Figure reports the distribution of missing R&D cells within firms over 1977 to 2022 across the firms in our Compustat sample.

Figure A.7 reports the aggregate reported R&D flows across firms with no missing R&D cells as well as those with missing cells accounting for less or more than 50 percent. As indicated in the figure,

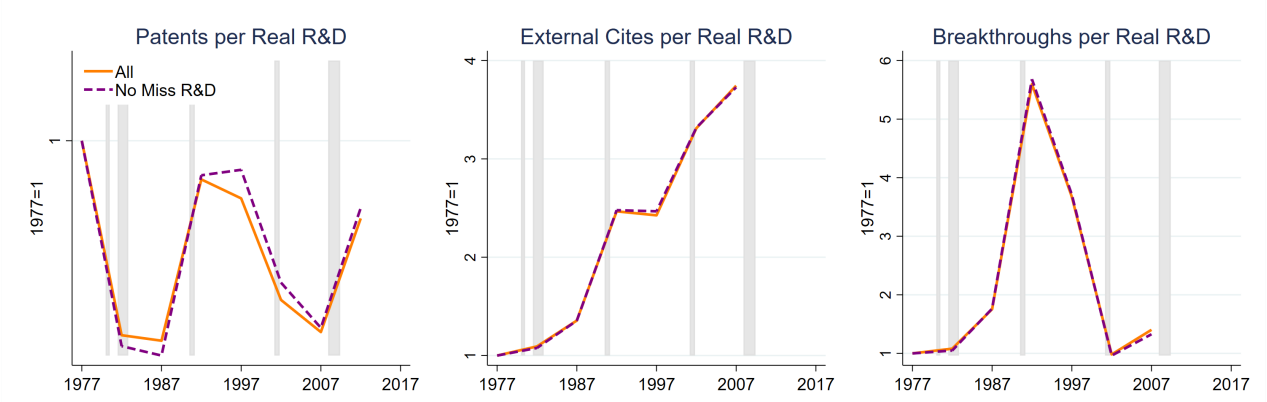
Figure A.7: Aggregate R&D Flow by Firm Type



Source: PV, KPSS, DS, and authors' calculations. Figure reports total real R&D expenditure across firms in our Compustat sample with no missing R&D cells as well as those with fewer or more than 50 percent of such cells from 1977 to 2022.

Figure A.8 examines the sensitivity of the simple efficiencies for Compustat firms displayed in Figure 4 of the main text to the inclusion of firms that have missing R&D cells. (In principle, if reporting increases over time, this could bias measured simple efficiency downwards in later years versus earliers years.).

Figure A.8: Compustat "Simple" research productivity Robustness to Firms with Missing R&D



Source: PV, LBD, KPST, RADS, BDSPF-Long, DS, Compustat and authors' calculations. Figure reports patents, breakthroughs, and external citations per real R&D expenditure for the firms tracked by the Compustat panel. Ratios are averages across 5-year semi-decades from 1977 to 2012, and these averages are indexed to 1 in the first, 1977 to 1981 semi-decade. Patents are assigned to firm by application year. Figure compares the trends for our baseline results in the main text to the those which exclude firms that have any missing R&D expenditures over the years they are present in Compustat. In both cases, results are restricted to firms that show up in our patent elasticity regressions. The orange lines in this figure are identical to the orange lines in Figure 4.

Table A.9: Patents Accounted for by Non-Missing Research Input

	LBD Panel				Compustat Panel	
Semi-Decade	All	5455	5417	RADS	R&D	SG&A
1970	100	77	54	65	94	98
1980	100	77	55	64	95	95
1990	100	75	48	59	96	91
2000	100	77	49	59	95	95
2010	100	77	48	59	94	96

Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, RADS, Compustat and authors' calculations. Table reports percent of patents accounted for by firm-year observations with non-missing research inputs. For the LBD panel, research inputs are: total pay, pay at Professional Services and Management (NAICS 54-55) establishments, pay at R&D labs, and R&D reported in the RADS survey. For the Compustat panel, research inputs are R&D expenditure in the Census RADS survey.

D Census R&D Surveys

The Census Bureau conducts R&D surveys, collectively referred to as RADS, with available data from 1972 to 2021. The surveys are conducted in collaboration with the National Center for Science and Engineering Statistics (NCSES) and used to publish aggregate US R&D statistics.[14]

The RADS comprise 4 different surveys that have changed over time. The Survey of Industrial

[14]These totals are publicly available at https://ncses.nsf.gov/data-collections/national-patterns/2021#data.

Research & Development (SIRD) was conducted from 1953 to 2007, followed by the Business Research & Development and Innovation Survey (BRDIS) from 2008 to 2016, the Business Research & Development Survey (BRDS) from 2017 to 2018, and the Business Enterprise Research & Development (BERD) survey from 2018 to the present. These surveys provide detailed information on firms' basic and applied R&D expenditures, as well as their funding source, *i.e.*, federal versus private firm.

The SIRD and BRDIS covered firms with five or more employees, while the BERD covered businesses with 10 or more employees. The BERD excludes companies that performed for funded less than $50k of R&D . While such firms were included in the past surveys, Census estimates suggest that such firms accounted for small portions of aggregate R&D. Starting in 2017, the Annual Business Survey (ABS) collected R&D activity for firms with one to nine employees. The ABS also collects information on business innovation activities.

The R&D surveys are collected by Employer Identification Number (EIN). We use the BR to collapse EINs to the firm level and develop code to the match the data to our LBD panel. In practice, the enterprise identifiers in the RADs are not always EINs. We cycle over various Census identifiers and also match to a window of years to improve matching. This procedure is especially important in certain years around transitions in the identifiers used in the underlying Census data.

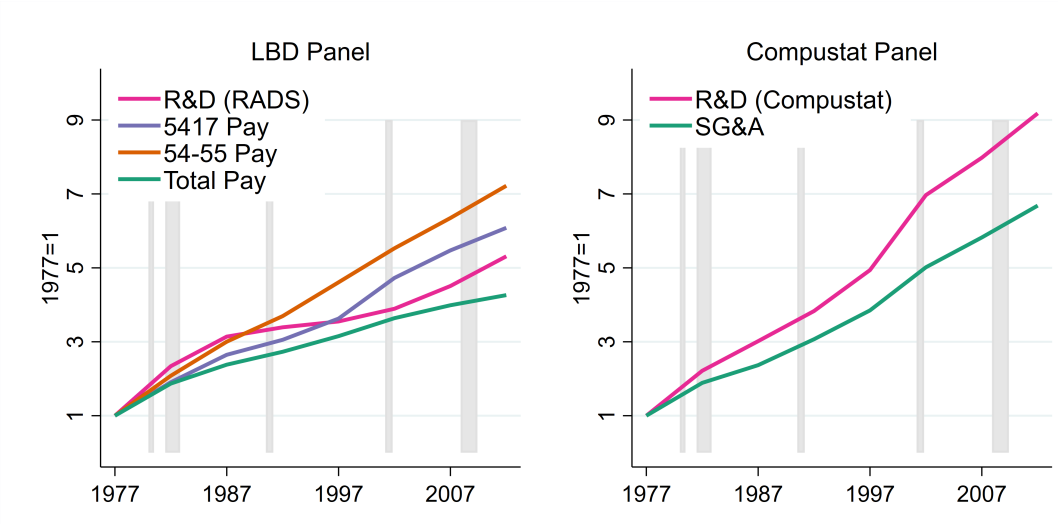
The survey re-designs were driven by changes in how R&D is performed and funded, and included changes to both the information collected and sampling frames. From 1957 to 1992, the SIRD focuses exclusively on manufacturing firms known to have conducted R&D in the previous 5 years. In 1992, sampling expanded substantially to include firms with unknown R&D expenditures. These "uncertainty" cases comprise the majority of the current sampling in terms of firms. The sampling frame is partitioned into three groups: companies known to perform R&D, companies that report two consecutive years of R&D expenditure; and companies whose R&D activity is unknown. To be included in the sample, a firm must meet a minimum size threshold (5 or more employees for the SIRD and 10 or more for the subsequent surveys). Firms that are in the top 50 of their state by payroll, or have an R&D lab (NAICS 5417) are included in the sample with certainty. Firms that report two consecutive years of zero R&D expenditures are dropped from future surveys. Surveyed firms are sent a "long" form if they are known R&D performers, and a "short" form otherwise. The "short" firm is designed to establish whether the firm engages in R&D, and thus does not contain many of the detailed questions appearing on the long form, which often requires different parts of a firm to weigh in on responses.

E Research-Input Stocks

We transform these research input flows into stocks using the perpetual inventory method, assuming a discount rate of 15 percent, consistent with prior literature (Griliches, 1998). We compute the payroll-based research inputs only for the years in which the firm has the relevant establishment(s), *i.e.*, we do not impute initial stock for firms with the relevant establishments in the first year of the LBD (1976), nor do we include any "residual" stock that depreciates if the firm closes the relevant establishments. We make this assumption based on the premise that the establishment 'houses' a

particular type of knowledge or expertise via its workers, which persists only while that establishment is alive and has such an industry code. When constructing research-input stocks based on the RADS, we assume flows are zero during any gaps in firms' participation in the survey and set the resulting stocks to missing in those years. We only use stocks constructed from unimputed flows, namely, we drop stocks after gaps in firms' participation in the RADS.[15] We follow a similar approach to compute SG&A and R&D real expenditure stocks for Compustat firms, though we treat missing flows as true zeros for this subset of publicly traded firms.

Figure A.9: Research-Input Stocks



Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, RADS, Compustat and authors' calculations.
Notes: Figure reports average annual real aggregate research-input stocks across firms in the LBD (left panel) and Compustat (right panel) by semi-decade. In the LBD panel, research inputs are firms' capitalized RADS R&D expenditure, R&D lab (NAICS 5417) payroll, Professional Services and Management (NAICS 54-55) payroll, and total payroll. Series are normalized to 1 in the first semi-decade. Reporting is done by semi-decade to minimize disclosure concerns. A data point for 2012, for example, is the average stock across firms from 2012 to 2016.

F Patent Types Over time

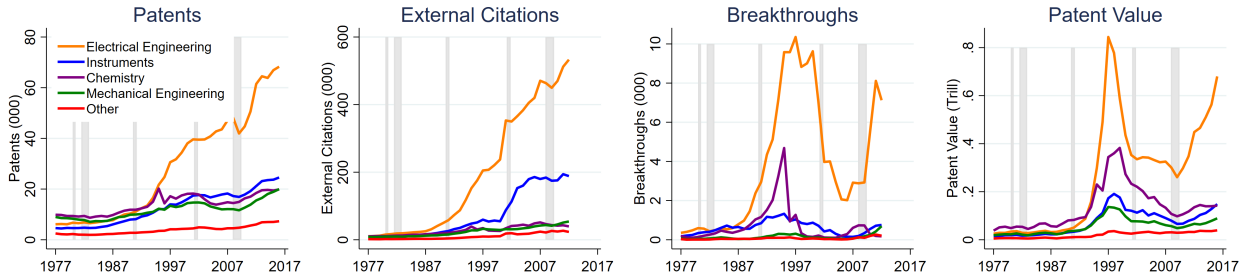
Figure A.10 presents a breakdown of US patenting activity by World Intellectual Property Organization (WIPO) categories, as recorded in PV.[16] The first panel of the figure shows patent counts, followed by counts of external citations in the second panel, counts of breakthrough patents in the third, and patent value in the fourth. In the first panel we see that the increase in patenting shown in Figure 2 is largely driven by innovation in Electrical Engineering, which includes patents related to

[15] In undisclosed results, we have computed two variants of research-input stocks. The first, applicable only to RADS R&D expenditure, imputes gaps in firms' survey responses with the average flow of the two endpoints before computing stocks. The second computes an initial stock for firms present in the first year of the LBD equal to $\bar{x}_f/(d+g)$ where $\bar{x}_f$ is the firm's average observed flow, and d and g are the depreciation and long-run flow growth rates, which we set to 15 and 2 percent, respectively.

[16] While Figure A.10 plots just the first WIPO category listed for each patent in PV, we find similar patterns when patents are assigned to all categories listed for them.

computing, telecommunications, information technology, and semi-conductors. This growth occurs in two bursts starting in the late 1980s and after the Great Recession.[17]

Figure A.10: Patenting Activity by WIPO Code



Source: PV, KPSS, KPST and authors' calculations. Notes: Figure provides a breakdown of patents, external citations, breakthrough patents and patent value by WIPO category and year. Patent value is in trillions of 1982 dollars.

In the second, third, and fourth panels, we see that Electrical Engineering patents also account for the greatest share of external citations, breakthroughs and patent value. There is less agreement among across these measures for the second largest contributor to patenting activity. Indeed, while Instrument patents are responsible for a surging share of external citations in the latter part of the sample period, there is no commensurate jump in Instrument breakthroughs or value. By contrast, patents in Chemistry, which includes pharmaceuticals and biotech, account for a relatively large share of value and breakthroughs *vis à vis* citations.

[17]The top 25 patents by value in the KPSS dataset are for Electrical Engineering. All but 3 are granted in the mid-to-late 1990s; all but 7 are classified as breakthroughs; and all but 6 were assigned to Cisco or Oracle.

G Additional Patent-by-Firm Statistics

Table A.10: Patents Per Firm, by Decade

	LBD Panel				Compustat Panel	
Semi-Decade	All	5455	5417	RADS	R&D	SG&A
1977	.22	1	4.3	4.3	.4	1.2
1982	.17	.7	3.2	4	.3	.8
1987	.18	.7	2.7	4.7	.2	.9
1992	.23	.8	2.7	5.4	.3	1.7
1997	.31	1	2.7	8.4	.4	2.4
2002	.32	1	2.5	6.5	.7	1.7
2007	.34	.9	2.3	4.9	1	1.5
2012	.48	1.3	2.9	5.4	1.6	2.1

Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, RADS, Compustat and authors' calculations. Table reports average annual patents per firm by semi-decade for four groups of firms: those that patent at least once in our sample, and the subsets of those firms that have a Professional Services and Management (NAICS 54-55) establishment, an R&D lab, or report R&D expenditure in the Census RADS survey.

Table A.11: Firm Counts, by Decade

	LBD Panel				Compustat Panel	
Semi-Decade	All	5455	5417	RADS	R&D	SG&A
1977	138	23	4	5	14	4
1982	172	31	5	5	18	6
1987	207	42	8	5	21	9
1992	235	51	10	7	26	10
1997	261	63	14	6	25	10
2002	269	71	17	10	20	8
2007	269	76	20	14	18	7
2012	262	76	20	16	17	6

Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, RADS, Compustat and authors' calculations. Figure reports firm-year counts (in thousands) by semi-decade for four groups of firms: those that patent at least once in our sample, and the subsets of those firms that have a Professional Services and Management (NAICS 54-55) establishment, an R&D lab, or report R&D expenditure in the Census RADS survey.

Table A.12: Breakdown of Patents, Citations and Breakthroughs by Decade (LBD Panel)

	Patents					External Citations					Breakthroughs				
Sector	1970	1980	1990	2000	2010	1970	1980	1990	2000	2010	1970	1980	1990	2000	2010
30 Manufacturing	78	72	61	42	30	77	69	54	33	23	73	58	42	16	12
42 Wholesale	3	5	10	17	15	4	5	12	18	18	6	7	16	22	15
51 Information	3	4	3	8	12	4	5	5	13	15	7	8	9	26	37
54 Professional	2	4	8	14	19	3	5	10	16	18	3	9	11	16	19
55 Management	5	8	9	9	11	4	7	10	9	12	3	5	12	13	10
Other	8	8	9	10	14	8	9	9	12	13	8	13	11	7	8
Size															
1–19	7	7	9	9	10	8	8	10	10	11	6	7	6	6	5
20–99	7	8	8	8	8	7	9	9	9	9	5	9	6	5	5
100–499	7	7	8	8	8	6	8	8	8	8	4	7	7	5	4
500–2499	8	8	10	11	10	8	8	10	10	10	6	7	8	8	8
2500–9999	13	11	15	15	16	13	12	16	15	14	10	10	16	12	11
10,000+	58	58	49	49	48	58	55	47	48	49	70	60	57	65	67
Age															
0–5 years	2	6	8	8	7	2	9	11	10	10	2	9	8	8	5
6–10 years	0	2	7	6	6	0	3	8	8	7	0	4	8	7	5
11–15 years	0	0	6	6	6	0	0	8	7	7	0	0	9	8	14
16–20 years	0	0	4	6	7	0	0	5	7	8	0	0	7	7	10
21–25 years	0	0	1	6	6	0	0	1	6	6	0	0	1	5	6
26+ years	0	0	0	4	13	0	0	0	5	13	0	0	0	3	12
Left Censored	98	91	74	64	55	98	87	67	57	49	98	87	68	62	48

Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, RADS, Compustat and authors' calculations. Figure provides a breakdown, in percent, of patent grants, external citations and breakthroughs by LBD firms' major 2-digit NAICS sector, size, and age. Firms in the LBD cannot be observed prior to 1977. As a result, the majority of firm observations have left-censored age in the early years of the panel. Cells representing less than 0.5 percent are coded as 0.

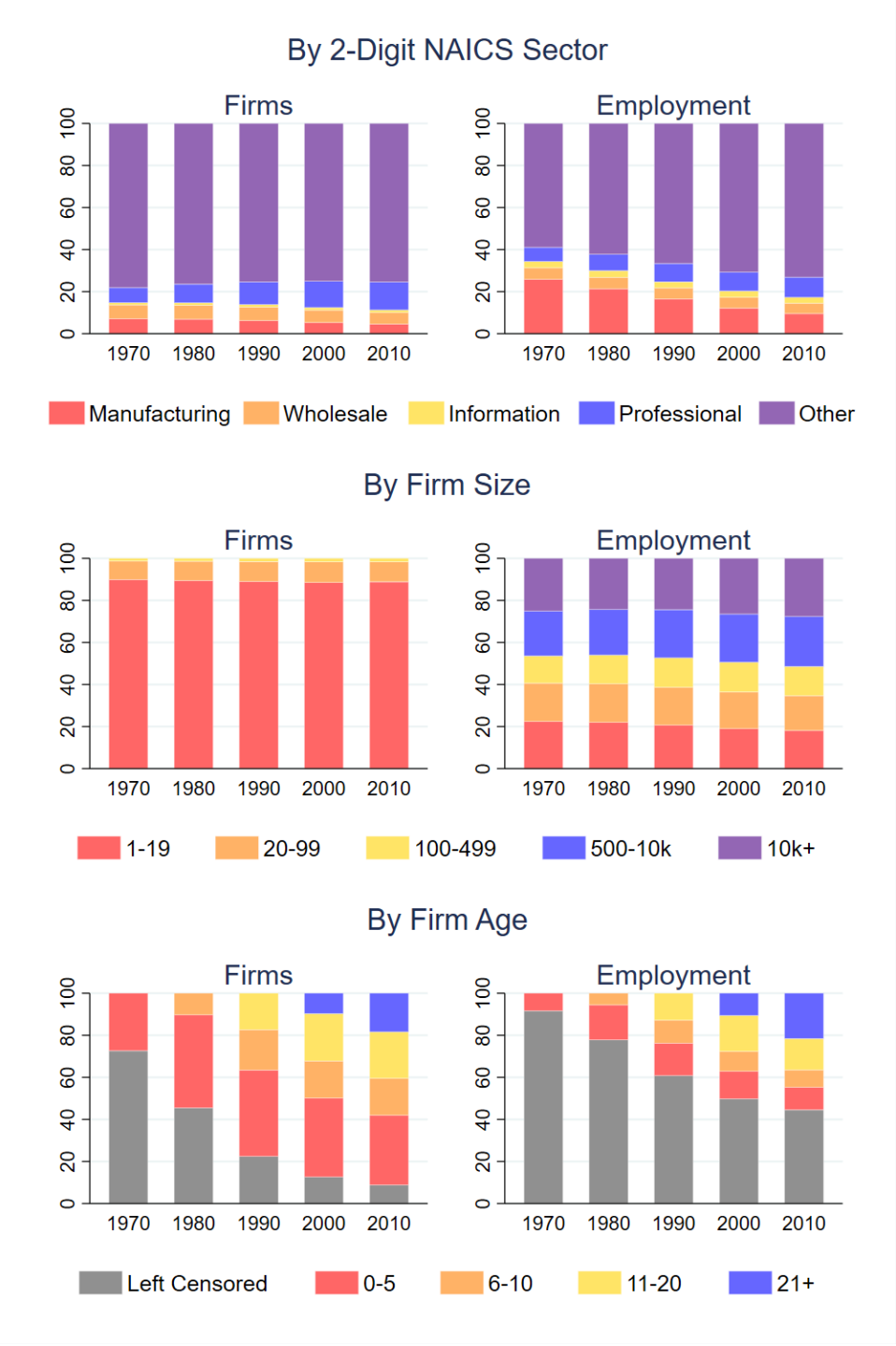
Table A.13: Breakdown of Patents, Citations and Breakthroughs by Decade (Compustat Panel)

	Patents					External Citations					Breakthroughs				
Sector	1970	1980	990	2000	2010	1970	1980	990	2000	2010	1970	1980	990	2000	2010
30 Manufacturing	71	78	78	70	62	73	81	80	68	66	64	67	61	40	37
42 Wholesale	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
51 Information	8	9	12	20	25	7	7	10	19	20	23	23	31	45	51
54 Professional	1	1	3	1	1	1	1	3	2	2	4	2	5	3	2
Other	20	12	8	9	12	19	10	7	11	13	9	8	4	11	11
Size															
1–19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20–99	0	1	2	1	1	0	1	3	1	1	0	2	1	0	0
100–499	1	2	4	4	3	1	3	7	7	5	1	5	5	3	2
500–9999	13	13	18	20	17	15	16	26	28	26	8	11	17	16	12
10,000+	86	84	76	76	80	83	79	65	63	68	90	82	77	80	86
Age															
0–5 years	0	4	12	11	8	1	7	18	16	12	0	8	13	11	8
6–10 years		1	7	9	7		2	11	13	8		3	10	11	15
11–20 years		0	10	19	16		0	14	21	23		0	16	28	20
21+ years			0	9	24			0	12	25			0	10	28
Left Censored	100	95	71	52	45	99	91	57	38	32	100	89	60	40	29

Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, RADS, Compustat and authors' calculations. Figure provides a breakdown, in percent, of patent grants, external citations and breakthroughs by LBD firms' major 2-digit NAICS sector, size, and age. Firms' sectors are time-invariant. Firm age represents years present in Compustat, and is computed here as if firms cannot be observed prior to 1977 to maintain consistency with Figure 5. Cells representing less than 0.5 percent are coded as 0.

Figure A.11 depicts the number of firms and workers by firm sector, size, and age. These figures are made from the US Census Bureau's public Business Dynamics Statistics (BDS).

Figure A.11: Firms and Employment by Firm Sector, Size and Age (BDS)



Source: BDS and authors' calculations.
Notes: Figure provides a breakdown of US firms and employment by firms' major 2-digit NAICS sector, size, and age as reported in Census Bureau's publicly available Business Dynamics Statistics (BDS) database available at https://www.census.gov/programs-surveys/bds.html. Bars representing less than 0.5 percent of patents are suppressed.

Table A.14: Firm Counts and Employment by Sector, Size and Age (LBD Panel)

	Firms					Employment				
Sector	1970	1980	1990	2000	2010	1970	1980	1990	2000	2010
30 Manufacturing	7	7	6	5	5	26	21	17	12	10
42 Wholesale	7	7	6	6	5	5	5	5	5	5
51 Information	1	1	1	1	1	3	3	3	3	3
54 Professional	7	9	11	13	13	7	8	9	9	10
Other	78	76	75	75	75	59	62	67	71	73
Size										
1–19	90	89	89	88	88	22	22	21	19	18
20–99	9	9	9	10	9	18	18	18	17	17
100–499	1	1	2	2	2	13	14	14	14	14
500–9999	0	0	0	0	0	21	22	23	23	24
10,000+	0	0	0	0	0	25	24	24	27	28
Age										
0–5 years	27	44	41	38	33	8	16	15	13	11
6–10 years	0	10	19	18	18	0	6	11	9	8
11–20 years	0	1	17	23	22	0	1	13	17	15
21+ years	0	0	0	10	18	0	0	1	11	22
Left Censored	73	45	22	13	9	92	77	60	50	45

Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, RADS, Compustat and authors' calculations. Figure provides a breakdown, in percent, of patent grants, external citations and breakthroughs by LBD firms' major 2-digit NAICS sector, size, and age. Firms in the LBD cannot be observed prior to 1977. As a result, the majority of firm observations have left-censored age in the early years of the panel. Cells representing less than 0.5 percent are coded as 0.

H Patent Elasticity Estimates

Table A.15 presents the patent elasticity estimates for the universe of patenting firms depicted in the top panels of Figures 8 and 9. Table A.16 presents the estimates for the Compustat sample depicted in the bottom panels of those figures.

Table A.15: Patent Elasticity Estimates

	Patents				Breakthroughs				External Cites				Patent Value			
	Total	5455	5417	RADS	Total	5455	5417	RADS	Total	5455	5417	RADS	Total	5455	5417	RADS
j=1982	-.5662***	-.7315***	-.6818**	-.809***	.6816***	.3311	1.59*	-.5139	.1565	-.1968	-.05548	-.4648**	.3087	.1933	.9138	.5076
	(.1242)	(.1542)	(.3062)	(.1694)	(.244)	(.375)	(.8326)	(.5671)	(.1606)	(.2391)	(.4342)	(.2025)	(.8696)	(.8096)	(1.109)	(.9575)
j=1987	-.5524***	-.7096***	-.01644	–1.245***	.7816***	.7766	2.679***	–1.085*	.5743***	.3964	.8656*	-.7196**	1.66	1.176	2.058*	1.391
	(.1505)	(.2526)	(.4322)	(.3027)	(.2847)	(.5056)	(.8862)	(.6463)	(.1973)	(.3475)	(.5032)	(.2918)	(1.419)	(1.21)	(1.235)	(1.168)
j=1992	-.3856*	-.1765	-.03553	–1.278***	1.06***	1.863***	3.372***	-.6897	1.124***	1.573**	1.533***	-.1619	4.903***	4.54***	2.855**	2.074
	(.2137)	(.4292)	(.4152)	(.4056)	(.3674)	(.6919)	(.9173)	(.7518)	(.3133)	(.6271)	(.5427)	(.6327)	(1.596)	(1.32)	(1.245)	(1.278)
j=1997	-.5837*	-.03857	.3345	–1.674**	–1.304**	.1976	2.107	–1.844	1.054***	1.777**	1.796**	-.5353	5.544***	6.059***	4.027***	3.757**
	(.3072)	(.6315)	(.6228)	(.6919)	(.5675)	(.9362)	(1.371)	(1.681)	(.409)	(.7852)	(.7777)	(1.026)	(1.862)	(1.567)	(1.232)	(1.572)
j=2002	–1.29***	-.6836	.2423	–2.155***	–4.124***	–2.08***	1.278	–3.32*	.8046**	1.793***	2.798***	-.06887	5.434***	6.93***	5.186***	4.747***
	(.3164)	(.5793)	(.6442)	(.7324)	(.6774)	(.7181)	(1.388)	(1.788)	(.3634)	(.594)	(.8266)	(.914)	(2.012)	(1.628)	(1.273)	(1.559)
j=2007	–1.851***	–1.133	-.3119	–2.732***	–4.661***	–1.247	1.227	–3.185	.1257	1.419**	2.658***	-.2668	5.034**	7.041***	5.231***	5.259***
	(.3898)	(.7061)	(.8473)	(.9899)	(.7021)	(.8403)	(1.299)	(1.989)	(.4049)	(.6232)	(.9613)	(1.039)	(1.992)	(1.815)	(1.316)	(1.532)
j=2012	–2.222***	–1.403*	-.6483	–3.333***									5.468***	7.864***	5.724***	5.31***
	(.458)	(.8116)	(.9981)	(1.14)									(2.09)	(2.11)	(1.308)	(1.644)
Stock x j=1977	.4649***	.2601***	.08763**	.265***	.4722***	.2101***	.1426**	.1609	.4135***	.2688***	.1509***	.3146***	.6118***	.4458***	.1443	.2888***
	(.0343)	(.0394)	(.0403)	(.07356)	(.05152)	(.06015)	(.06791)	(.1022)	(.03673)	(.05491)	(.04793)	(.08659)	(.1179)	(.1017)	(.1006)	(.1057)
Stock x j=1982	.4754***	.2914***	.1288***	.2992***	.426***	.1935***	.04061	.1957*	.4011***	.2848***	.1677***	.3444***	.5702***	.4126***	.06941	.2411***
	(.03509)	(.04116)	(.0432)	(.07489)	(.05425)	(.06822)	(.08643)	(.1159)	(.03707)	(.05516)	(.06077)	(.08651)	(.09458)	(.08057)	(.04884)	(.08398)
Stock x j=1987	.4812***	.297***	.0953**	.3347***	.4509***	.199***	.00967	.2662**	.3953***	.2655***	.1301**	.3797***	.4921***	.3495***	-.003259	.1875**
	(.03481)	(.03794)	(.0462)	(.07514)	(.05196)	(.04952)	(.06636)	(.1034)	(.03456)	(.04208)	(.05411)	(.07868)	(.08407)	(.07526)	(.04519)	(.07588)
Stock x j=1992	.4852***	.2738***	.1138***	.3516***	.4962***	.1939***	.03422	.3054***	.3987***	.2239***	.1257***	.3812***	.3531***	.1867**	-.003341	.1976***
	(.03019)	(.0331)	(.03572)	(.05513)	(.05207)	(.05055)	(.0647)	(.09233)	(.03301)	(.03929)	(.04463)	(.05845)	(.08483)	(.07322)	(.03811)	(.07252)
Stock x j=1997	.5116***	.2792***	.1064**	.3941***	.6173***	.2859***	.105	.3633***	.4193***	.2293***	.1299***	.4233***	.3525***	.131	-.03368	.1371
	(.02832)	(.04111)	(.04491)	(.0479)	(.05743)	(.06253)	(.09883)	(.1365)	(.03351)	(.04426)	(.04836)	(.06334)	(.1103)	(.09524)	(.0349)	(.1009)
Stock x j=2002	.5555***	.3252***	.1152**	.4256***	.7098***	.3576***	.07774	.3738***	.4572***	.256***	.0829**	.4152***	.3219***	.03827	-.1557***	.03734
	(.03025)	(.03727)	(.04534)	(.0497)	(.07728)	(.06274)	(.08684)	(.1438)	(.03764)	(.03849)	(.03803)	(.06615)	(.1217)	(.0979)	(.04007)	(.0828)
Stock x j=2007	.5895***	.3556***	.1563***	.4592***	.7588***	.3277***	.09796	.3847**	.5038***	.2892***	.09792**	.4291***	.3262***	.01362	-.1886***	-.01917
	(.02989)	(.04342)	(.05882)	(.0467)	(.07642)	(.07312)	(.07319)	(.1569)	(.04157)	(.04763)	(.04255)	(.07253)	(.1234)	(.1193)	(.04491)	(.08832)
Stock x j=2012	.6241***	.3864***	.1981***	.5063***									.3191**	-.01632	-.2015***	-.004054
	(.03277)	(.05082)	(.07185)	(.05322)									(.1371)	(.1476)	(.05528)	(.106)
Constant	–2.699***	.9297**	3.74***	1.187	–4.734***	-.4515	.8871	.5904	-.8275*	1.873***	3.946***	1.643	–2.28	1.021	5.566***	3.263**
	(.4262)	(.4708)	(.4534)	(.935)	(.6879)	(.7382)	(.8203)	(1.37)	(.46)	(.6634)	(.5988)	(1.117)	(1.78)	(1.35)	(1.181)	(1.429)
Constant	–2.699***	.9297**	3.74***	1.187	–4.734***	-.4515	.8871	.5904	–2.28	1.021	5.566***	3.263**	-.8275*	1.873***	3.946***	1.643
	(.4262)	(.4708)	(.4534)	(.935)	(.6879)	(.7382)	(.8203)	(1.37)	(1.78)	(1.35)	(1.181)	(1.429)	(.46)	(.6634)	(.5988)	(1.117)
R-Square	.8519	.8894	.9126	.8916	.7876	.8157	.8433	.8310	.907	.8994	.9112	.8929	.8835	.9053	.924	.9088
Observations	1813000	430000	96500	68000	186000	78000	24500	23500	86500	53500	22000	29000	1290000	310000	70500	50000
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Source: LBD, PV, KPST, KPSS and authors' calculations. Table reports estimates of equation 8. Each column focuses on a different patent outcome and research-input stock. First set of coefficients report estimates of the semi-decade fixed effects displayed in Figure 9. Second set of estimates report the semi-decade varying patent elasticities displayed in Figure 8.

Table A.16: Patent Elasticity Estimates (Compustat Panel)

	Patents		Breakthroughs		External Cites		Patent Value	
	SG&A	R&D	SG&A	R&D	SG&A	R&D	SG&A	R&D
j=1982	–0.917***	–0.701***	0.0758	0.0493	–0.277**	–0.262***	–0.689**	–0.277
	(0.138)	(0.0919)	(0.325)	(0.264)	(0.134)	(0.0900)	(0.298)	(0.329)
j=1987	–1.263***	–0.848***	–0.350	–0.262	–0.0588	–0.0592	–0.802**	0.0418
	(0.239)	(0.162)	(0.451)	(0.423)	(0.192)	(0.143)	(0.389)	(0.676)
j=1992	–0.572	–0.343	0.780	0.872*	1.155***	0.949***	1.578***	1.361**
	(0.362)	(0.211)	(0.539)	(0.509)	(0.260)	(0.172)	(0.453)	(0.679)
j=1997	–0.162	–0.163	–0.151	0.269	1.470***	1.157***	2.893***	2.282***
	(0.505)	(0.296)	(0.727)	(0.802)	(0.331)	(0.228)	(0.503)	(0.597)
j=2002	–0.217	–0.0881	–2.153**	–0.919	1.559***	1.471***	2.645***	1.946***
	(0.610)	(0.393)	(1.030)	(1.293)	(0.394)	(0.297)	(0.515)	(0.437)
j=2007	–0.417	–0.106	–2.279*	–0.912	1.449***	1.460***	2.874***	2.057***
	(0.743)	(0.472)	(1.186)	(1.471)	(0.481)	(0.361)	(0.481)	(0.399)
j=2012	–0.528	–0.146					3.394***	2.372***
	(0.852)	(0.616)					(0.528)	(0.396)
Stock x j=1977	0.359***	0.370***	0.455***	0.384***	0.401***	0.414***	0.356***	0.457***
	(0.0651)	(0.0564)	(0.0862)	(0.0811)	(0.0454)	(0.0407)	(0.0577)	(0.0955)
Stock x j=1982	0.437***	0.426***	0.437***	0.363***	0.437***	0.448***	0.424***	0.462***
	(0.0674)	(0.0566)	(0.0886)	(0.0893)	(0.0448)	(0.0391)	(0.0612)	(0.0574)
Stock x j=1987	0.492***	0.462***	0.549***	0.487***	0.444***	0.456***	0.455***	0.435***
	(0.0712)	(0.0572)	(0.0926)	(0.0915)	(0.0438)	(0.0370)	(0.0509)	(0.0513)
Stock x j=1992	0.434***	0.418***	0.532***	0.474***	0.354***	0.375***	0.272***	0.356***
	(0.0745)	(0.0539)	(0.0914)	(0.0922)	(0.0463)	(0.0368)	(0.0474)	(0.0515)
Stock x j=1997	0.422***	0.442***	0.611***	0.533***	0.348***	0.384***	0.188***	0.313***
	(0.0850)	(0.0601)	(0.105)	(0.125)	(0.0523)	(0.0422)	(0.0503)	(0.0473)
Stock x j=2002	0.429***	0.428***	0.688***	0.517***	0.391***	0.398***	0.163***	0.283***
	(0.0976)	(0.0733)	(0.149)	(0.191)	(0.0573)	(0.0474)	(0.0503)	(0.0483)
Stock x j=2007	0.443***	0.419***	0.717***	0.545***	0.408***	0.402***	0.105**	0.218***
	(0.105)	(0.0780)	(0.157)	(0.203)	(0.0647)	(0.0543)	(0.0448)	(0.0454)
Stock x j=2012	0.476***	0.451***					0.0925*	0.223***
	(0.115)	(0.0963)					(0.0494)	(0.0691)
Constant	2.073***	2.718***	–0.612	0.542	2.316***	3.012***	4.033***	4.197***
	(0.393)	(0.230)	(0.569)	(0.389)	(0.258)	(0.180)	(0.423)	(0.549)
Observations	96,233	74,516	36,300	33,322	83,134	64,843	41,328	38,922
R-Square	0.908	0.904	0.838	0.831	0.907	0.904	0.920	0.922
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y

Source: Compustat, PV, KPSS, KPST and authors' calculations. Table reports estimates of equation 8. Each column focuses on a different patent outcome and research-input stock. First set of coefficients report estimates of the semi-decade fixed effects displayed in Figure 9. Second set of estimates report the semi-decade varying patent elasticities displayed in Figure 8.

I Growth Regression Estimates

Table A.17 presents the coefficient estimates depicted in Figures 13 and 14.

Table A.17: Growth Regression Estimates (LBD Panel)

Dependent variable:	Growth(Sales/Worker)			Growth(Sales)		
Idea Stock:	Patents	Breakthroughs	External Cites	Patents	Breakthroughs	External Cites
Growth(Employment)				.8741***	.8665***	.8758***
				(.01007)	(.0185)	(.01223)
j=1982	.2053***	.2284	.1945**	.22***	.216***	.2236***
	(.07421)	(.1573)	(.08953)	(.03907)	(.07659)	(.03979)
j=1987	.4248***	.5583***	.4273***	.3529***	.3744***	.3396***
	(.0746)	(.1551)	(.0883)	(.03788)	(.07219)	(.03966)
j=1992	.203***	.3727**	.2407***	.2289***	.2815***	.2366***
	(.07165)	(.1482)	(.08611)	(.03311)	(.06019)	(.03451)
j=1997	.09297	.2018	.1122	.2094***	.2255***	.2073***
	(.07181)	(.1479)	(.08688)	(.03182)	(.05816)	(.03332)
j=2002	.1953***	.3831***	.2163**	.2498***	.2868***	.2523***
	(.07279)	(.1466)	(.08817)	(.03355)	(.06294)	(.03633)
j=2007	.08921	.2132	.1202	.1471***	.1585***	.1536***
	(.06984)	(.1456)	(.08494)	(.03086)	(.05859)	(.03245)
j=2012	.09846			.152***		
	(.06905)			(.03058)		
Growth(Stock) x j=1977	.1235***	-.02259	.1235***	.01727	-.04346	.03207*
	(.03764)	(.0307)	(.03324)	(.01812)	(.02693)	(.01636)
Growth(Stock) x j=1982	.47***	.2173***	.428***	.08682***	.03265	.0829***
	(.03569)	(.05916)	(.03458)	(.01398)	(.02511)	(.0144)
Growth(Stock) x j=1987	.2903***	.1876***	.2786***	.04779**	.089*	.06735***
	(.03757)	(.06076)	(.03808)	(.02253)	(.04641)	(.02575)
Growth(Stock) x j=1992	.2281***	.1013***	.2058***	.02293*	.003478	.02028
	(.02749)	(.03186)	(.02592)	(.01225)	(.01822)	(.01264)
Growth(Stock) x j=1997	.439***	.3837***	.4069***	.1046***	.1172***	.1024***
	(.02695)	(.0454)	(.02792)	(.02002)	(.04388)	(.02607)
Growth(Stock) x j=2002	.3498***	.4209***	.3327***	.07165***	.09644***	.068***
	(.03288)	(.06487)	(.0308)	(.01536)	(.02941)	(.01471)
Growth(Stock) x j=2007	.3344***	.2722***	.3424***	.04066***	.03589*	.04089***
	(.02959)	(.04741)	(.02935)	(.01402)	(.02131)	(.01349)
Growth(Stock) x j=2012	.365***			.0638***		
	(.02647)			(.01206)		
Constant	-.297***	-.3516***	-.3373***	-.1662***	-.1755***	-.1729***
	(.06239)	(.1274)	(.0746)	(.02668)	(.04926)	(.0276)
R-Square	.1949	.1666	.1913	.6920	.6543	.6896
Observations	321,000	33,000	226,000	409,000	42,000	291,000
Firm FE	N	N	N	N	N	N
Weighted	Y	Y	Y	Y	Y	Y

Source: LBD, PV, KPSS, KPST and authors' calculations. Table reports estimates of equation 10 via OLS. Each column focuses on a different combination of growth and idea stock. First set of coefficients report estimates of the semi-decade fixed effects displayed in Figure 9. Second set of estimates report the semi-decade varying patent elasticities displayed in Figure 8. Regressions also include firm size and age bins, whose coefficients are not reported.

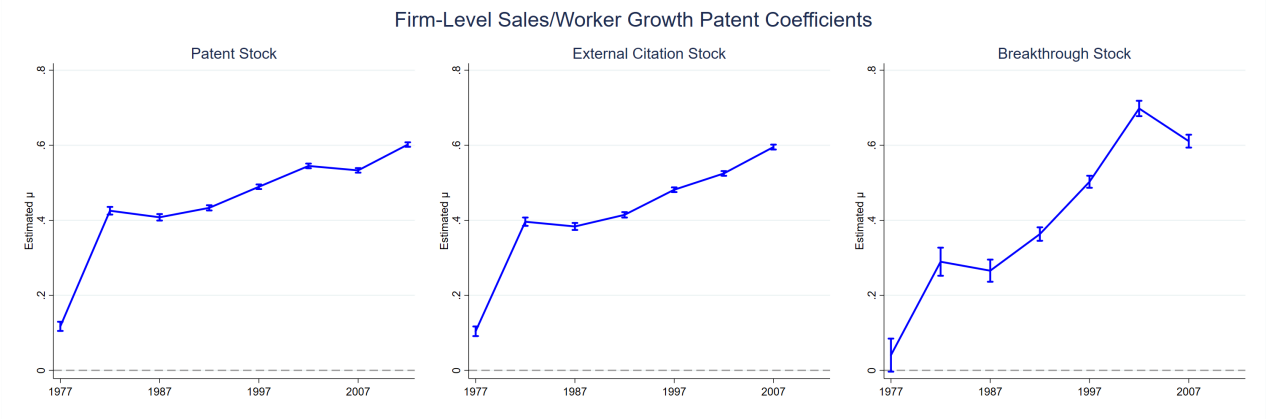
Table A.18: Growth Regression Estimates (Compustat Panel)

Dependent variable:	Growth(Sales/Worker)			Growth(Sales)		
Idea Stock:	Patents	Breakthroughs	External Cites	Patents	Breakthroughs	External Cites
Growth(Employment)				0.515***	0.545***	0.587***
				(0.0716)	(0.0739)	(0.0862)
j=1982	0.149***	0.166***	0.0398	0.192***	0.0397	–0.0717
	(0.0302)	(0.0348)	(0.0277)	(0.0490)	(0.0350)	(0.0399)
j=1987	0.385**	0.457**	0.214	0.470***	0.383**	0.294**
	(0.118)	(0.149)	(0.129)	(0.0909)	(0.126)	(0.111)
j=1992	0.438***	0.506***	0.266***	0.348***	0.278**	0.167
	(0.0580)	(0.0857)	(0.0694)	(0.0540)	(0.0818)	(0.0889)
j=1997	0.311***	0.336***	0.187**	0.278***	0.239**	0.120
	(0.0725)	(0.0844)	(0.0672)	(0.0523)	(0.0756)	(0.0723)
j=2002	0.427***	0.492***	0.279***	0.531***	0.525***	0.304***
	(0.0534)	(0.0946)	(0.0507)	(0.0452)	(0.0823)	(0.0780)
j=2007	0.0860	0.143	–0.0493	0.177**	0.185*	–0.0305
	(0.0604)	(0.0972)	(0.0720)	(0.0676)	(0.0930)	(0.111)
j=2012	–0.0146	0.141	–0.0515	0.0587	0.256**	0.0103
	(0.0492)	(0.0859)	(0.0574)	(0.0675)	(0.0799)	(0.0748)
Stock x j=1977	0.0295	0.0487**	–0.0371*	0.160***	0.0774**	–0.0824***
	(0.0247)	(0.0190)	(0.0164)	(0.0326)	(0.0276)	(0.0143)
Stock x j=1982	0.197**	0.150***	0.164***	0.286***	0.305***	0.0866**
	(0.0603)	(0.0416)	(0.0444)	(0.0339)	(0.0390)	(0.0299)
Stock x j=1987	–0.0546	–0.161	0.0152	0.235***	0.229***	–0.0197
	(0.0590)	(0.0952)	(0.0546)	(0.0443)	(0.0552)	(0.0279)
Stock x j=1992	–0.0469	–0.100	0.128**	0.271***	0.269***	0.0830**
	(0.0584)	(0.0676)	(0.0484)	(0.0493)	(0.0579)	(0.0320)
Stock x j=1997	–0.0832	–0.0136	0.0530	0.204***	0.190***	0.167***
	(0.0519)	(0.0476)	(0.0385)	(0.0532)	(0.0421)	(0.0336)
Stock x j=2002	–0.00579	–0.0383	–0.0412	0.135	0.0251	0.0204
	(0.0502)	(0.0422)	(0.0362)	(0.0811)	(0.0577)	(0.0448)
Stock x j=2007	0.236***	0.200*	0.153***	0.432***	0.324**	0.133**
	(0.0590)	(0.0916)	(0.0336)	(0.0556)	(0.0955)	(0.0514)
Stock x j=2012	0.181**	0.394***	0.105*	0.492***	0.587***	0.156**
	(0.0585)	(0.0709)	(0.0546)	(0.0599)	(0.0761)	(0.0641)
Constant	0.0196	–0.0529	0.230	–0.0665	–0.119	0.206
	(0.181)	(0.226)	(0.204)	(0.228)	(0.266)	(0.303)
Observations	3,197	3,196	2,249	3,201	3,200	2,253
R-squared	0.176	0.187	0.207	0.523	0.520	0.506
Firm FE	N	N	N	N	N	N
Weighted	Y	Y	Y	Y	Y	Y

Source: LBD, PV, KPSS, KPST and authors' calculations. Table reports estimates of equation 10 via OLS. Each column focuses on a different combination of growth and idea stock. First set of coefficients report estimates of the semi-decade fixed effects displayed in Figure 9. Second set of estimates report the semi-decade varying patent elasticities displayed in Figure 8. Regressions also include firm size and age bins, whose coefficients are not reported.

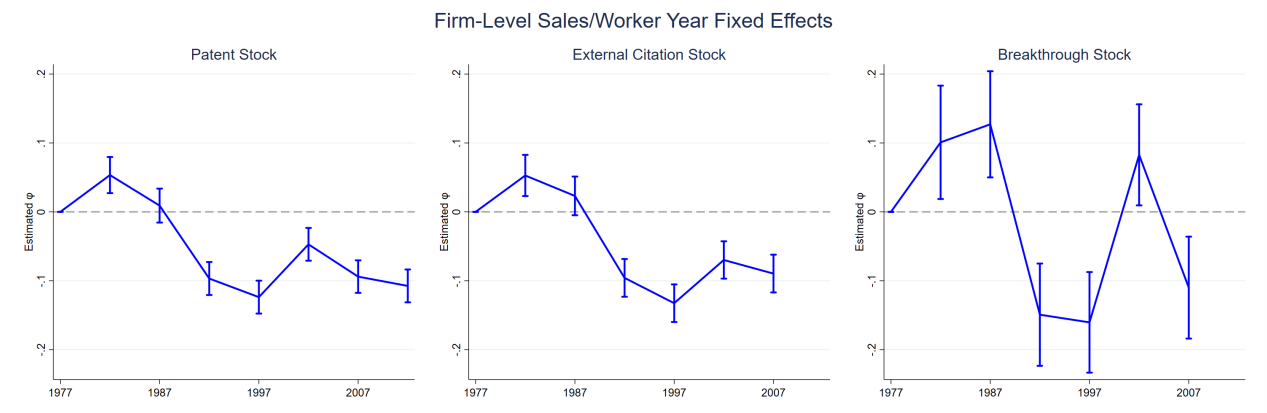
I.1 Unweighted Growth Regression Estimates

Figure A.12: Unweighted Sales Per Worker Growth Coefficient Estimates (LBD Panel)



Source: LBD, EC, PV, KPST, and authors' calculations. Notes: Figure reports relationship between LBD firms' sales per worker and idea stock growth between years t and t+5. These estimates correspond to unweighted estimates of μ_j in equation (10). Regressions include firm size and age bin controls and year fixed effects. Standard errors are clustered at the firm-level with 95% CIs depicted.

Figure A.13: Estimated Time Fixed Effects in Unweighted Sales per Worker Growth Regression (LBD Panel)



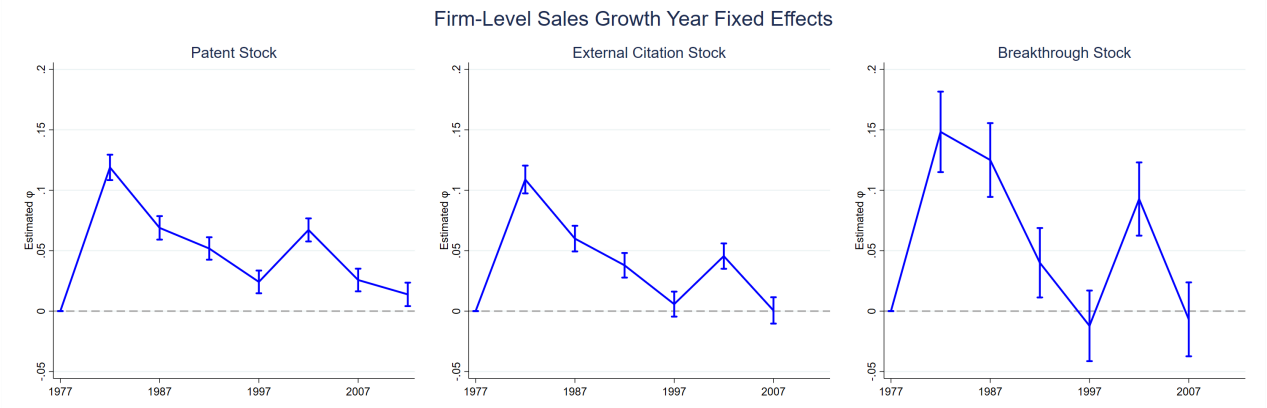
Source: LBD, EC, PV, KPST, and authors' calculations. Notes: Figure reports estimated time fixed effects from regressing LBD firms' sales on their idea stock growth between years t and t+5. These time fixed effects are from unweighted estimates of ϕ_j in equation (10). Regressions include firm size and age bin controls and year fixed effects. Standard errors are clustered at the firm-level with 95% CIs depicted.

Figure A.14: Unweighted Sales Growth Coefficient Estimates (LBD Panel)



Source: LBD, EC, PV, KPST, and authors' calculations. Notes: Figure reports relationship between LBD firms' sales and idea stock growth between years t and t+5. These estimates correspond to unweighted estimates of μ_j in equation (10), while controlling for the firm's employment growth. Standard errors are clustered at the firm-level with 95% CIs depicted.

Figure A.15: Estimated Time Fixed Effects in Unweighted Sales Growth Regression (LBD Panel)



Source: LBD, EC, PV, KPST, and authors' calculations. Notes: Figure reports estimated time fixed effects from regressing LBD firms' sales on their idea stock growth between years t and t+5. These time fixed effects are from unweighted estimates of ϕ_j in equation (10), while controlling for the firm's employment growth. Standard errors are clustered at the firm-level with 95% CIs depicted.