Growth is Getting Harder to Find, Not Ideas*

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

April 10, 2025

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 46-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. Over our sample period, we find that researchers' patenting productivity is increasing, there is little evidence of any secular decline in high-quality patenting common to all firms, and the link between patents and growth is present, differs by type of idea, and is fairly stable. On the other hand, we find strong evidence of secular decreases in output unrelated to patenting, suggesting an important role for other factors. Together, these results invite renewed empirical and theoretical attention to the impact of ideas on growth. To that end, our patent-firm bridge, which will be available to researchers with approved access, is used to produce new, public-use statistics on the Business Dynamics of Patenting Firms (BDS-PF).

*We thank participants at the Fall 2024 Yale Trade Day and UC Davis trade seminar, and Samuel Kortum, Bitsy Perlman, and Cheryl Grim for helpful comments and discussions. 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.
§Yale Department of Economics.
¶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 innovation. In his Nobel-prize-winning research into endogenous growth, Romer (1990) shows that firms' decisions to invest in research and development can affect the economy's long-run growth rate, even when the population is constant. A key assumption supporting this result is that the number of ideas researchers discover each year rises linearly with the stock of ideas already discovered. If, instead, researchers' marginal products tail off as the knowledge stock increases, firms would increasingly reallocate workers to production, and growth would decline. Romer noted the need for empirical evidence to justify this assumption: "Whether opportunities in research are actually petering out, or will eventually do so, is an empirical question that this kind of theory cannot resolve."[1]

Subsequent examination of US data reveals dramatic increases in both the number and share of workers devoted to research, even as the aggregate US growth rate has remained fairly constant. These patterns have led some researchers to conclude that "ideas are getting harder to find" (Jones, 1995; Kortum, 1997), since output growth maps one-to-one with idea growth in Romer (1990). Most recently, Bloom et al. (2020) expand on this evidence by documenting similar patterns for various samples of products and firms.[2] They posit that research efforts may face diminishing returns, perhaps because all of the "low hanging fruit" has been picked. The ensuing "semi-endogenous" growth models assume that researchers' marginal products exhibit decreasing returns in the aggregate stock of ideas, such that the growth rate declines as research output grows (holding the number of researchers fixed). In these models, innovation affects average consumption per person, but has no impact on long-run growth rates. As such, there is no role for innovation policy to affect long-run growth; improvements in living standards depend solely on population growth.

While a constant aggregate output growth rate coupled with the rising share of US researchers is consistent with the hypothesis that a growing stock of knowledge makes it harder to innovate, existing papers do not provide direct evidence of such a mechanism. Moreover they do not pursue other possible explanations for the aggregate trends. Most notably, this interpretation neglects the possibility that the relationship between ideas and growth, as opposed to that between research inputs and ideas, is not linear, may be influenced by other factors, and can change over time. Even if ideas abound, mapping them into growth may depend on the type and quality of ideas, the extent to which they generate winner-take-all rewards or foster increases in market power, changing incentives firms face to implement ideas, and evolution of regulatory and competitive environments.

In this paper, we build a novel 46-year firm-level panel to examine these two relationships separately – between knowledge inputs and ideas, and between ideas and growth. Toward that end, we construct a new bridge mapping the universe of US non-farm employer businesses in the Census

[1] Romer (1990), page S84. Related papers in which endogenous growth arises from directed research efforts yield similar predictions. For example, see Grossman and Helpman (1991) and Aghion and Howitt (1992).

[2] For example, they compare R&D expenditures by firms that produce semiconductors to real output growth in semiconductors, R&D expenditures in US agriculture and changes in crop yields, the number of cancer research publications versus years of life saved after a cancer diagnosis, and R&D expenditures versus output growth by a subset of Compustat firms and a subset of US manufacturing firms.

Bureau's Longitudinal Business Database (LBD) to US patents, which, following the literature, we take as a key manifestation of ideas. This bridge is used to produce new, public-use Business Dynamics of Patenting Firms (BDS-PF) statistics available on Census' website. To address the well-known limitations of simple patent counts, our analysis also considers patents weighted by external citations (Akcigit and Kerr, 2018), novelty (Kelly et al., 2021), and private monetary value (Kogan et al., 2017). Exploiting detailed information on firms' payroll and R&D expenditures we construct a broad set of knowledge-input stocks, and estimate *within-firm* elasticities of ideas to knowledge, as well as the relationship between the flow of ideas and firm-level sales growth. Crucially, we allow both sets of estimates to vary over time.

We start by documenting three new facts about the evolution of US patenting that have important implications for future work. First, it is no longer sufficient to use the manufacturing sector to study US innovation. 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 the most novel patents (i.e., "breakthroughs") falling from 73 to 12 percent. Second, patenting differs markedly across the firm size and age distributions, with considerable changes in both over time. 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. Patenting by young firms rises over time, and young firms account for disproportionately high shares of external citations and breakthroughs relative to their share of patent grants. Finally, our comparison of trends in our LBD panel versus a similarly constructed Compustat panel reveals that researchers relying on the latter largely miss not only cross-sectional differences in the types of firms that patent, but also how patenting firms' characteristics evolve over time.

The changing patterns of patenting firms' industry, size, and age present a range of potential explanations for why researcher productivity and the effects of ideas on growth may change over time. For instance, innovation by manufacturing firms may be sold in more tangible goods with changes in value that are easier to measure. Young and small firms in nascent industries may be better able to create new ideas, but less well-equipped to bring those ideas to market. Large and old firms, which are better able to scale ideas, may also have incentives to exploit market power to restrain output, thereby obscuring the relationship between research inputs and growth in aggregate data.

In the second part of the paper, we estimate potentially time-varying relationships between firm-level patenting and our various measures of knowledge-input stocks. In contrast to the notion that "ideas are getting harder to find," the estimated patent elasticities are flat or *rising* over time with respect to almost all the knowledge inputs we consider. For the most broadly defined knowledge input – capitalized total payroll – we find that the elasticity for patents rises from 0.46 in the first semi-decade of our analysis (1977 to 1981) to 0.62 in the final semi-decade (2012 to 2016). For our most narrowly defined knowledge stock – capitalized R&D expenditures – elasticities increase from 0.26 to 0.48 over the same interval. These results suggest that firms' ability to discover new ideas, per unit of knowledge inputs, has improved over the last five decades.

Ideas may differ in terms of their quality and thus likelihood of affecting growth. To determine whether *high-quality* ideas also are getting easier to find, we estimate analogous elasticities using patents weighted by the external citations they receive within five years of granting, and by whether or not they are a breakthrough. We also consider patents' estimated market value, which contains investors' assessments of the contribution of the idea to firms' net present value. This indicator of patent quality is unlike citations and breakthroughs because it incorporates the second of the two key relationships mapping research inputs to growth, i.e., the market's expectation about the link between the idea and growth. We find that while patent elasticities are similarly flat or rising over time for patents weighted by external citations and breakthroughs, they *fall* for value. The first two results indicate that one definition of "ideas are getting harder to find," i.e., declining ideas per research input controlling for other factors, is not evident in the data over the last five decades. The third result suggests that the mapping of those ideas to sales or productivity growth may be changing.

An important consideration in estimating elasticities over our long sample period is that the process of patenting might change for various reasons, e.g., in terms of the quality of the ideas firms seek to protect, or the way in which the US Patent and Trademark Office (USPTO) grants approval. A key feature of our estimation is that we include both firm and semi-decade fixed effects. The latter capture changes in patenting across semi-decades that are common to all firms in a particular sample. This approach ensures that our patenting elasticities are not contaminated by common institutional or macro-economic shocks that might raise or lower the simplest measure of patent efficiency – patents per researcher – often examined in the growth literature.

The estimated semi-decade fixed effects are also informative in their own right. Indeed, if, in contrast to Romer (1990)'s assumption, a rising stock of ideas somehow makes finding new ideas more difficult, one would expect a secular decline in these fixed effects. While the data reveal such declines for two of our six specifications using patent grants, the semi-decade fixed effects from the citation-weighted grants and patent value regressions are flat or rising across all measures of knowledge stocks, while those for breakthrough-weighted patents are flat or rising for the first half of our sample period and decline thereafter. These trends do not support a second way of defining "ideas are getting harder to find," i.e., that idea creation has fallen universally across all firms independent of research efforts over the last five decades.

In the final section of the paper, we examine the second link between ideas and growth directly. We follow the growth regressions introduced by Kogan et al. (2017), but extend them in three ways: first to allow the semi-elasticity between growth and ideas to evolve over time, second to examine additional patenting activities, and third to follow the growth literature more closely by regressing the growth of output on the growth of ideas, controlling for changes in employment.[3] We generally find a positive relationship between the growth rate in a firm's patent grants and its sales growth, even after controlling for changes in its employment growth. Growth is more strongly related to citations and breakthrough patents, relative to patent grants. Most notably, the estimates are fairly stable over time. By contrast, we observe declines at the end of the period in the semi-decade fixed effects after controlling for firms' growth of ideas. These decreases in the time effects suggest a secular decline in growth among patenting firms that is independent of their growth in ideas.

[3]In ongoing work for a future draft, we are replicating these regressions using the LBD panel.

Together, our results indicate that researcher productivity is rising, there is no clear trend of a secular decline in firms' ability to find ideas, and that, if anything, the link between ideas and growth is rising. The trends are hard to square with the claim that "ideas are getting harder to find." On the other hand, we find that the relationship between growth and ideas varies by the type of idea, and document a secular decline in output growth after controlling for changes in firms' ideas. These results suggest greater attention be paid to the link between ideas and growth in trying to understand why the rise in researcher inputs has not translated to an increase in growth rates.

A central contribution of this paper is to develop and share a new set of algorithms to match the USPTO patent data to the US Census data. This bridge will be available to all researchers with approved projects through the Federal Statistical Research Data Centers (FSRDCs) and is used to produce the Census Bureau's public-use Business Dynamics Statistics of Patenting Firms (BDS-PF) tables. Our bridge provides the longest period of matched data, spanning 1976 to 2021[4], and matches 92 percent of US-based patent-assignee records. To improve matches, and (partially) account for spurious changes in firm identifiers over time, we develop a new method to ensure longitudinal consistency in our matching algorithms that exploits the persistence of a particular firm in the USPTO data. To our knowledge, we are the first to develop and share these new longitudinal matching techniques, which can be generalized to improve the matching of other external datasets to the Census Bureau's micro data.

We also contribute to three strands of the literature. First, we add to the research that estimates the elasticity of patents to R&D expenditure. A seminal paper in this body of work, Hausman et al. (1984), develops a Poisson specification to estimate patent elasticities, with more recent contributions by Howell (2017) and Meyers and Lanahan (2022). Our contribution is to use the Poisson Pseudo Maximum Likelihood (PPML) estimator from Silva and Tenreyro (2006) to estimate patent elasticities by semi-decade for the universe of US firms and with respect to a wide range of patent measures and R&D inputs. We construct a novel 46-year firm-level panel that encompasses the universe of US non-farm employer firms and use it to show these elasticities are flat or rising for patent grants and patent grants weighted by external citations and novelty, and declining for patent value.

We also add to work studying the impact of innovation on firm-level productivity and growth. Griliches (1979) introduces the notion of an innovation production function, and a long line of work in industrial organization estimates the effect of R&D expenditure on firm productivity (Hall et al., 1998; Peters et al., 2017). Past work also links a firm's first successful patent grant to increases in its size and profits (Balasubramanian and Sivadasan, 2011; Kline et al., 2019). Kogan et al. (2017) use stock market valuation changes to estimate the value of publicly traded firms' granted patents and show that this value is a strong predictor of future growth. We confirm that firms' patent value predicts future growth, but also find that citation-weighted patent counts and breakthrough patents predict growth. Most notably, we provide the first evidence that this relationship between patents and growth has been changing over time.

Finally we contribute to a growing body of work that investigates the empirical support for

[4]Our analysis data begin in 1977 while the data match, available to FSRDC projects with approved projects, begins in 1976.

whether ideas are getting harder to find. This work relates to a broader debate about long-run trends in economic growth and whether all of the "low-hanging fruit" technologies have been picked (Jones, 2009; Cowen, 2011; Brynjolfsson and McAfee, 2014; Gordon, 2017). 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). Ekerdt and Wu (2024) argue that declining researcher productivity is expected in any setting with heterogeneous workers who select into production versus innovation based on their relative skills, since inframarginal innovation workers will necessarily be less productive. Bloom et al. (2020) infer a decline in researcher productivity by comparing total growth in output to total growth in R&D expenditures across various samples of aggregated micro data. Our contribution is to estimate both the relationship between innovation inputs and ideas, as well as the relationship between ideas and growth and allowing each to change over time. While a long-line of macro work assumes that constant growth rates in light of increased R&D expenditures must signify concavity in Romer (1990)'s innovation production function, our findings instead point to the need for a new way to model the production of final goods and services from ideas.

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 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 Linking Knowledge Inputs to Growth

Romer (1990) presents a model in which firms' profit-maximizing decisions to invest in idea creation influence long-run growth. In his setting, the aggregate number of 'designs' (A) is a nonrival input to final-good production and evolves according to

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

where $\dot{A}$ is the change in the stock of designs, H_A is total human capital (or researchers) devoted to innovation, and δ measures what Romer calls 'researcher productivity'. Final output is produced by a representative firm using the following Cobb-Douglas technology

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

in which each intermediate good x_i requires an idea i (or blueprint) to produce, L is labor, and H_Y is human capital devoted to final-good production. Although there is an infinite potential number of ideas, only those that have been discovered can be used to produce. The knowledge stock thus determines the number of inputs in the economy. Since inputs have additively separable effects on output, there is a one-to-one mapping between $\dot{A}$ and output growth. Population is fixed by assumption, so growth depends on the share of workers in the population that perform research,

$H_A/(H_A + H_Y)$, and their productivity.

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 over time while aggregate growth has remained fairly constant. Jones (1995) proposes a 'semi-endogenous' growth model in which $\dot{A}$ need not be linear in A, e.g.,

$$\dot{A} = \delta H_A A^{1-\beta} \tag{3}$$

as a way to reconcile Romer (1990) with this macro fact. In this formulation, $\beta > 0$ implies that the growth rate in ideas will decline over time as the existing stock of knowledge expands (Jones, 1995; Kortum, 1997).[5]

Although the formulation in equation (3) is one way to reconcile equation (1) with aggregate trends, at least two alternative explanations are worthy of attention. First, researcher productivity (δ) might vary over time for a variety of reasons, including policy. To the best of our knowledge, none of the empirical research in the growth literature ties the inferred decline in researcher productivity to increases in the *aggregate* stock of ideas (A) versus changes in research productivity (δ). A second possibility is that the relationship between ideas and output may be changing. As noted above, Romer (1990) assumes that all ideas have additively separable effects on final-good output. In his words, "An investigation of complementarity as well as a of mixtures of types of substitutability is left for future work."

Empirical investigation of long-run growth must also confront the existence of firms. Romer (1990) assumes away any role for firm boundaries, with final-goods produced by a single competitive firm. In reality firms innovate in part to obtain monopoly rents over their competitors. Innovating firms' incentives to exploit market power and constrain output may therefore affect the extent to which aggregate output responds to aggregate innovation. Our descriptive results in Section 4 provide suggestive evidence that these factors may indeed be at play.

Our econometric analysis seeks to shed light on these issues by separately considering the relationship between knowledge inputs and ideas versus ideas and growth. For the former, we consider

$$Ideas_{ft} = K_{ft}^{\eta_t} \gamma_f \gamma_t \gamma_{ft}, \tag{4}$$

where η_t is the elasticity of ideas to firm f's knowledge stock (K_{ft}) in year t. γ_f captures a time-invariant firm ability to generate ideas, which might be due to a wide range of factors such as better research practices. γ_t accounts for annual variation in idea creation that is common to all firms. These time fixed effects capture the contribution of the aggregate stock of knowledge available to all firms, e.g., A in Equation (1). They may also reflect stochastic discovery of new technology "classes", e.g, the PC-internet revolution of the late 1980s and 1990s, or artificial intelligence in the late 2010s, in the spirit of Ribeiro (2024). To the extent that the 'crowing out of ideas' is a an economy-wide phenomenon that affects all firms equally and is increasing in the aggregate knowledge stock, it might

[5] Romer (1990) explicitly acknowledges 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.

manifest in a secular decline in these year effects. Finally, γ_{ft} represents time-varying idiosyncratic shocks to firm innovation.

To distinguish the creation of ideas from the production of output, our empirical analysis in Section 5 uses patent grants as a measure of ideas. Patent grants provide a simple metric to capture an idea that is sufficiently novel to merit legal protection. To identify high-quality ideas, we also use patent counts weighted by external citations and novelty. We use a broad range of knowledge-input stocks constructed from firms' payroll and R&D surveys to approximate K_{ft}.

In Section 6, we consider an analogous mapping to study the second relationship between ideas (patenting) and growth,

$$\frac{\Delta Y_{ft}}{Y_{ft}} = (Ideas_{ft})^{\mu_t} \varphi_f \varphi_t \varphi_{ft}, \tag{5}$$

where $\frac{\Delta Y_{ft}}{Y_{ft}}$ represents growth in firms' output, operating profit, or TFP, μ_t is the elasticity of growth to ideas, φ_f captures a time-invariant firm ability to convert ideas into growth due, e.g., to better management, φ_t captures annual US variation in this ability, common to all firms, and φ_{ft} represents idiosyncratic shocks.

In the next section, we describe the detailed microdata we use to estimate these relationships within the firm.

3 Data

A central contribution of this paper is to construct a new 46-year firm-level panel dataset of US patents and knowledge input stocks. In this section, we describe the US patent data, the microdata on US firms and establishments from the US Census Bureau, and the Compustat data on publicly traded firms. We then describe how we merge the patent data to the two firm-level panels, and provide a brief comparison of patent coverage in the Census versus Compustat data. Finally, we detail our new measures of knowledge input stocks and discuss their coverage and patterns.

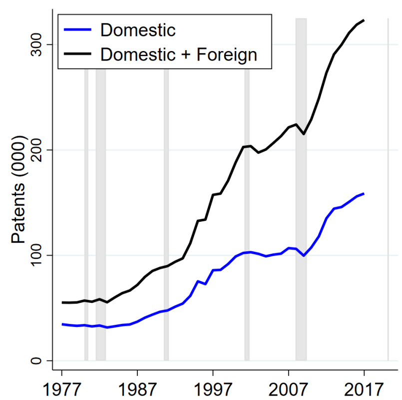
3.1 Patent Data

PatentsView (PV): The US Patent and Trademark Office (USPTO) provides information on the identity and location of all granted US patents and their corresponding assignees and inventors, dates of application and granting, and citations of other patents via their PatentsView portal. We currently observe patent grants and data on US firms (discussed below) through 2022. 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. Thus, about 90 percent of patents applied for in 2016 have been granted by 2022. As a result, 2016 is the current upper limit on the sample periods examined in our empirical analyses.

Figure 1 reports the relative prevalence of foreign and domestic patent grants by application year in PatentsView. Growth in both patent types is broadly similar, with two periods of sharp growth beginning in the 1980s and after the Great Recession. Domestic patents represent a declining share of all US patents, from about two-thirds in 1977 to about half in 2017, with most of the divergence occurring during the 2000s.[6]

We restrict the analyses in the remainder of the paper to "domestic" patents, which we define as patents with at least one assignee located in the United States; we treat all other patents as "foreign." We focus on domestic patents for two reasons. From a practical perspective, we have no US geographic information on purely foreign patents, and the LBD contains no information on firms or establishments without US operations. From a conceptual perspective, we would potentially overstate researcher patent efficiency by including foreign inventions in their idea counts.[7] Hereafter, all references to patents refer to domestic patents, unless otherwise specified.

Figure 1: Domestic vs Foreign Patents by Application Year



Source: PV and authors' calculations.
Notes: Figure reports total (black line) and domestic (blue 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 several limitations to using patent statistics as a measure of innovative activity. First, not all inventions are patented, either because certain types of inventions are not patentable, or innovating firms wish to avoid disclosing their innovations (Cohen et al., 2000). Second, patents differ greatly in their quality or economic impact (Bessen, 2008; Kogan et al., 2017). Since this paper relies on patents as a measure of innovative output, our analysis may underestimate the extent to which knowledge inputs generate 'ideas'. To address potential variation in patent quality and impact, we rely on two commonly used measures of patent quality: "external" citations received and patent novelty. We also use the stock market's predicted patent value, though note that this measure includes both a quality and an expected growth component. We describe each greater detail below.

External Citations: We use data on patent citations from PV to compute the total number of

[6] Appendix Figure A.7 provides a similar foreign versus domestic breakdown for patents weighted by each of the quality measures we consider below.

[7] Only a subset of foreign patents can be matched to firms in the United States since not all foreign patents are associated with firms that have employment in the United States. Moreover, foreign patents may 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.

citations received by each patent from other patents within the first five years of being granted. Windowing citations in this way accounts for the fact that older patents have more time to accumulate citations. We also distinguish between citations received from a firm's own subsequent patents ("self" citations) versus those from other firms' patents ("external" citations). Self-citation generally indicates "internal innovation," i.e., extending one's existing technologies in-house (Galasso and Simcoe, 2011), while external citation can capture exploratory innovation (Akcigit and Kerr, 2018). We therefore focus on 5-year external citations as a measure of patent quality, and include citations from both granted patents and pre-grant publications. In classifying citations, we consider a citation from a patent for which we do not have a matched firm identifier (about 8 percent of patents) to be an external citation. Since citations are forward-looking, this measure is available only through 2011.[8]

KPST Breakthrough Patents: Kelly et al. (2021) – hereafter KPST – estimate patent "novelty" as the ratio of patents' forward textual similarity to backward textual dissimilartiy with other patents over intervals of 5 or 10 years. Their updated dataset considers over 11.4 million patents granted from 1836 to 2022, and defines 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. In our analysis, we define breakthrough patents according to the 5-year window and 5-percent threshold. As breakthroughs, like citations, are forward-looking, this measure is also available only through 2011.

An important advantage of KPST's estimate of patent novelty is its longitudinal consistency. Kuhn et al. (2020), for example, argue that the nature of patent citations has changed substantially in recent decades, potentially complicating the use of citations as a measure of quality.[9] The KPST measure avoids this issue by using the same algorithm to analyze all patents' text. A second advantage of the KPST measure is its high correlation with productivity growth, which the authors document in their analysis.

KPSS Patent Value: Kogan et al. (2017) – hereafter KPSS – provide a measure of a patent's monetary value by estimating the abnormal returns in a firm's stock market price on the day it receives the patent grant. 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. Since these values are estimated using public stock data, the number of patents for which this measure is available is substantially lower than that for external citations and breakthroughs. On the other hand, because estimation relies only on data contemporaneous to granting, it is available in our sample through 2016.

Patent value is distinct from our other measures of ideas because it contains the market's best

[8] As noted at the beginning of this section, our sample is constructed from patents with application years up to 2016 that were granted by 2021. For some measures we require a five-year forward window, for example for citations received. Therefore, forward-looking measures are only observed for patents with application years up to 2011.

[9] Kuhn et al. (2020) shows that the technological similarity of citing and cited patents falls over time, indicating a change in the information content of a citation, and that the vast majority of recent citations are made by a relatively small number of patents. Fadeev (2023) argues that external citations mostly capture explicit, voluntary transfers of knowledge between collaborating firms. Despite this evidence, results using external citations are similar to those using breakthroughs and patent novelty.

estimate of the expected returns for that particular idea. In this sense, it contains both of the relationships we seek to estimate in Equations (4) and (5). Estimated patent value also reflects the expected private returns to an idea, rather than its societal value. For instance, a patent representing an important scientific breakthrough that garners many 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. Indeed, Kogan et al. (2017) find that firms' productivity rises with their own patent value and falls with the patent value of their competitors.

3.2 US Census Establishment- and Firm-level Data

US Census Bureau Micro Data: We construct a firm-level panel of employment, payroll, sales, and R&D expenditure by combining multiple Census Bureau datasets. 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 2022 (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.[10] To remain consistent with the Business Dynamics Statistics, and retain only the highest quality records, we keep only those establishments that are in scope for the Business Dynamics Statistics data. We also explicitly drop establishments in Public Administration (NAICS 92) and Agriculture (NAICS 11), since these are out-of-scope for the Economic Censuses and thus we have no sales or other Census information for them.

We aggregate these 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 we observe every establishment's primary industry code. We use the vintage-consistent North American Industrial Classification (NAICS) codes developed by Fort and Klimek (2018) to track the mix of employment within a firm across sectors over time. We use this measure to assign firms to their primary sector in each year based on the majority of their payroll. A firm's sector may thus change over time as the firm's mix of establishments changes, or as its establishments change industries. We also exploit the detailed establishment-level industry information to construct firms' knowledge-input stocks, as described in more detail below. We follow the BDS to 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.

Finally, we merge this firm panel with firm-level R&D expenditures for the subset of firms in the National Science Foundation's R&D surveys, collectively referred to as RADS. These surveys provide information about domestic and foreign R&D expenditures on basic and applied research for a rotating sample of approximately 45 thousand relatively large firms each year. The surveys disproportionately target large manufacturing firms, although the sampling frame has expanded over time. Appendix D provides additional details on the R&D surveys and how we merge them to the

[10] 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.

firm-level panel.

3.3 Compustat Firm Panel

We construct a similar firm-level panel using Compustat data. Compustat, a product of S&P Global Market Intelligence, contains financial information for publicly traded firms derived from companies' SEC filings, including 10-K and 10-Q forms. Compustat contains a wide range of alternate 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, which are constructed from firms' public filings with the US Securities and Exchange Commission.

3.4 Matching Patents to Firm-level Panel Data

This paper develops a new set of algorithms to match the USPTO patent data to the US Census data that provide four contributions relative to past work. First, the bridge developed here, termed "BDS-PF Long," will be available to qualified researchers with approved projects through the FSRDCs. The new bridge is also used to produce business dynamics statistics for patenting firms, available publicly in the BDS-PF on Census' website.[11]

Second, the BDS-PF Long matches 92 percent of US-based patent assignee records with a match precision rate of about 94 percent (see Appendix Tables A.1 and A.6). We developed this match precision rate using the subset of "fully triangulated" matches from the existing BDS-PF Triangulation bridge, first developed by Graham et al. (2018) but available only from 2000 forward, treating that subset as 'truth.'[12] The precision match rate is the number of correct matches divided by the total number of matches (correct and incorrect). We use this rate, common in matching and machine learning classification literature, to quantify our match quality and to guide choices in designing our matching algorithms.

Third, the bridge provides the longest period of matched data, spanning 1977 to 2021. For context, Kerr and Fu (2008) match roughly 77 percent of patent-assignee records in the 1976 to 2000 period, versus 94 percent in our new BDS-PF Long for those years. The BDS-PF Triangulation bridge has a similar match rate as BDS-PF Long, approximately 91 percent of US-based patent-assignee records, but only provides matches starting in 2000. Appendix Figure A.1 depicts our match rates over time for both US and foreign patent assignees and compares them to these existing crosswalks.

Fourth, we develop a new method to enhance the longitudinal consistency of matches, exploiting the persistence of a particular firm in the USPTO data, which allows us to leverage information across years to improve the quality of matches. To our knowledge, we are the first to develop and share these new longitudinal matching techniques, which we believe can be applied more broadly to link other external datasets to the Census Bureau's micro data.

[11] For additional information about the FSRDCs, see `https://www.census.gov/about/adrm/fsrdc.html` and for information on the public-use BDS tabs see `https://www.census.gov/programs-surveys/bds.html`.

[12] The BDS-PF Triangulation bridge uses both firm name and geography matching, along with inventors' name and location and corresponding worker employment information from the Longitudinal Employer Household Database (LEHD) to match patents to firms.

Matching to Census Business Data: We combine the UPSTO patent data with the Census firm-level data via name and geography matching. We first collect the universe of employer businesses from the combined County Business Patterns and Business Register (CBPBR) microdata files. The CBPBR contains the universe of private non-farm employers in the United States, is constructed from administrative tax filing records, and provides the backbone for the LBD. Due to the relatively limited geographic detail available for patent assignees, we match patents at the firm level, though we exploit geographic information to do so.

To facilitate matching, we extract and standardize business name and geography information from both PatentsView and the CBPBR. After matching by different combinations of exact and fuzzy business name and geography, we utilize the PatentsView disambiguated patent-assignee identifiers to improve the longitudinal consistency of matched Census firm identifiers within a given PatentsView assignee over time. We match patents to firms by grant year, but for purposes of our analyses we associated patents to firms using *application* year so that they are most proximate in time to innovation investments. 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.

We highlight three key findings from our match. First, match rates decline over time, from 95 percent in early 1970s to about 90 percent in the late 2010s. This decline is also apparent in the Kerr and Fu (2008) crosswalk and is observed in the match between trademarks and firms developed by Dinlersoz et al. (2018). In Appendix Figure A.3, we show that the declining match rates are due primarily to the changing composition of the initial match quality to the CBPBR. As shown in Appendix Figure A.2, 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 (Block 4) both rise over time. Appendix A provides further details on the matching algorithms and match quality.

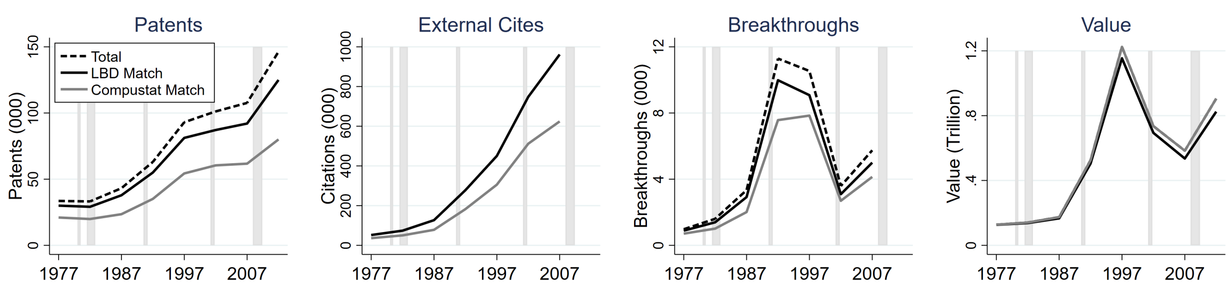
Matching Patents to Compustat: As a complement to our analyses using the Census Bureau's LBD, we create a separate match of US patents to the publicly traded firms tracked by Compustat. As discussed in greater detail in Appendix B, we construct a "meta-match" that combines two recent patent-to-firm mappings developed by KPSS and Dyevre and Seager (2024). We exploit two mappings to increase the overall number of patents we can match to Compustat, and to cross-validate patents encompassed by each approach. As with the LBD, we match domestic patents to firms by application year so that our matched patent is assigned as closely as possible to year in which its innovation occurred.

3.5 Patent Coverage in the LBD vs Compustat

Before continuing, we briefly summarize the four patenting measures introduced above and compare their coverage in the LBD versus Compustat panels. To minimize Census disclosure burden (here, and throughout the paper), we report averages across 5-year "semi-decades" anchored by quinquennial Economic Census years ending in "2" and "7". For example, a data point for 2012 is the average across 2012 to 2016.

The first panel of Figure 2 summarizes patenting, where the line for "Total" is total *domestic* patents in PatentsView, i.e., the trend for domestic patents displayed in Figure 1. Here, we see that the LBD match captures substantially more patents than the Compustat match, and that LBD coverage declines less steeply, from 89 to 85 percent of all patents in the first semi-decade, versus from 63 to 55 percent in the final semi-decade. The match rate for patents in our LBD panel is lower than the crosswalk match rate reported in Section 3.4 for several reasons. First, patents are matched to firms based upon proximity to grant year. However, our LBD analysis panel associated patents based upon application year. If we are unable to match a patent to a firm in the application year it will be unmatched in our LBD panel. Second, patents matched to entities outside our LBD panel, such as Government establishments or Universities, are unmatched in our LBD panel, even though we matched the patents to the Business Register. The remaining panels of Figure 2 summarize external citations, breakthrough patents, and patent value. We find that external citations rise more steadily throughout the sample period than patents in both the LBD and Compustat, with a sharper increase in the former than the latter.[13]

Figure 2: Patents Captured 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. External citations are defined only for matched patents, so no total is reported. 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; see main text for more detail. Legend for all panels is in the first panel. Patent value is in trillions of 1982 dollars and, as noted in the text, available for substantially fewer patents than the other series. All series are plotted by patents' application year.

By contrast, breakthroughs and patent value exhibit a pronounced inverted u-shaped pattern, increasing sharply in the early 1990s, coincident with the rise of the PC and internet, and then falling almost as precipitously in the early 2000s, though less sharply for value. Interestingly, the peak for breakthroughs occurs in the 1992 to 1996 period, whereas the peak for value occurs five years later. This gap could indicate that it takes time for firms to identify ways of appropriating the value of breakthrough technologies. Trends for the LBD and Compustat are similar, with the LBD panel capturing a larger share of breakthroughs, particularly at the peak. Coverage of patent value in the Compustat and LBD panels is almost identical, which is unsurprising given that it can be estimated only for publicly traded firms.[14]

[13]We use the Census *firmid* or Compustat *permno* to classify External Citations, so we do not have a "total" of unmatched external citations in the second panel.

[14]Coverage for patent value is slightly higher for Compustat firms at the end of the period, which suggests we succesfully matched some patents to a Compustat firm but not to a corresponding LBD firm. Such a match failure for public firm might occur, if for instance, the firm has mutiple identifiers in the LBD that we could not disambiguate.

3.6 Measuring Knowledge Inputs

Assessing whether ideas are getting harder to find requires measuring the resources firms devote to finding them, i.e., knowledge inputs. We first follow the literature and measure knowledge inputs using R&D expenditures. For the LBD panel, we use firms' total R&D expenditures as reported in the RADS surveys. For the Compustat panel, we measure knowledge inputs using SG&A and R&D real expenditures, which past work has used to measure firms' intangible capital (Corrado et al., 2005; Eisfeldt and Papanikolaou, 2014). In both cases, we deflate the nominal series using the Consumer Price Index (CPI).

The RADS cover a small and selected sample of firms that changes over time. To overcome this limitation, we use establishment payroll data to construct three measures of knowledge inputs based on firms' expenditures on likely R&D workers. To identify such workers we focus on establishments that are most likely involved in knowledge creation, namely those employed in R&D labs (NAICS 5417). While such labs are appealing due to their clear focus on innovation, they may miss considerable research efforts. We therefore construct a second measure based on the firm payroll in Professional, Scientific, and Technical Services as well as Management establishments (NAICS 54-55). Professional Services and Management (NAICS 54-55) includes establishments engaged in high-skill, more knowledge-intensive tasks such as engineering, computer system design.[15] 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. Finally, we provide an upper bound on firms' R&D employment expenditures using their total payroll across all establishments. We deflate all the payroll numbers using the CPI.[16]

We transform these knowledge 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 knowledge 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 knowledge 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. In our results going forward, we only use stocks constructed from unimputed flows, namely, we drop stocks after gaps in firms' participation in the RADS.[17] We

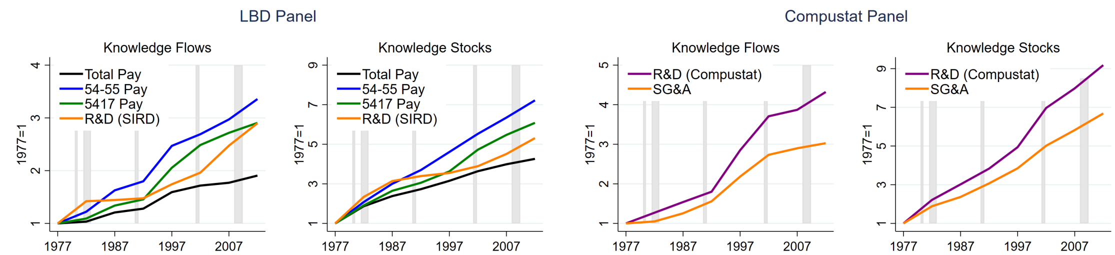
[15] NAICS 54 encompasses nine 4-digit NAICS industries: legal (5411), accounting (5412), engineering (5413), design (5414), computer system design (5415), consulting (5416), R&D (5417), advertising (5418), and other (5419).

[16] Our ability to count knowledge workers is limited by the fact that the LBD assigns all employment in an establishment to that establishment's primary NAICS code. Unfortunately, the Census data have limited (or no) information on worker occupations.

[17] In undisclosed results, we have computed two variants of knowledge 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.

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 3: Knowledge Inputs



Source: PV, LBD, KPSS, KPST, BDSPF-Long, DS, RADS, Compustat and authors' calculations.
Notes: Figure reports average annual real aggregate knowledge input flows and stocks across firms in the LBD (left panels) and Compustat (right panels) samples, by semi-decade. In the first two panels, knowledge inputs are firms' total, Professional Services and Management (NAICS 54-55) , and R&D lab (NAICS 5417) payroll, and RADS R&D expenditure. In the right two panels, they are total SG&A and R&D expenditures as recorded in Compustat. Series are normalized to 1 in the first semi-decade. Figure *f_kflow_compustat_title.png* generated in pks30.do.

For each input, the first two panels of Figure 3 report average aggregate flows and stocks across firms in the LBD by five-year semi-decades, normalized to be 1 for the 1977 to 1981 period. As indicated in the first panel, aggregate real total payroll flows nearly double between the first and last periods. Real Professional Services and Management (NAICS 54-55) and R&D lab (NAICS 5417) pay rise more quickly, each by a factor of about 3. Unsurprisingly, the path of R&D expenditure flows follow those of R&D lab (NAICS 5417) payroll fairly closely, both rising by 2.9 times between the first and last periods. However, during the mid- and late 1990s, R&D lab (NAICS 5417) payroll rises more quickly than R&D expenditures, then grows more slowly through the 2000s. The second panel reports analogous averages for stocks, which rise more quickly than flows given the magnitude of the flows and the constant 15 percent depreciation rate assumed in their calculation.

The right two panels of Figure 3 depict the SG&A and R&D real expenditure flows and stocks across all Compustat firms by semi-decade, again normalized to 1 in the first period. As indicated in the figure, Compustat SG&A flows and stocks are broadly comparable to those for Professional Services and Management (NAICS 54-55) payroll, while Compustat R&D flows and stocks are similar to those for RADS R&D expenditure in the LBD panel. In the Compustat panel, we see SG&A stocks rising by about 7 times and R&D expenditures rising by over 9 times between the first and last periods.

4 The Evolution of US Patenting Activity by Firm Type

In this section, we provide three new facts about the evolution of patenting firms in the United States over the last four decades. We first exploit the LBD panel to document significant shifts in the industry, size, and age distributions of patenting firms. We then demonstrate that these changes are largely missed when relying on the publicly traded firms captured in Compustat.

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. To summarize how these activities have changed over time, we provide 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.

Figure 4: Patenting Activity by Firm Sector, Size and Age (LBD Panel)

Source: PV, LBD, KPSS, KPST, BDSPF-Long, and authors' calculations. Figure provides a breakdown of patents, external citations, breakthrough patents, and patent value by LBD firms' major 2-digit NAICS sector, size and age. Bars representing less than 0.5 percent of patents are suppressed.

The top row of Figure 4 reveals a steady decline in manufacturing firms' share of patenting activity, which declines from 78 to 30 percent over the sample period. This decline is even starker for citations, breakthroughs, and value, which decrease from above 70 percent in the 1970s to 23, 12, and 18 percent, respectively, by the 2010s. By contrast, firms in Information (NAICS 51), Professional Services and Management (NAICS 54-55) and Wholesale (NAICS 42) sectors exhibit steady relative growth over time.[18] Indeed, firms in Professional Services and Management have larger activity shares than Manufacturing firms in the final decade across all four patent measures. Firms in Information are the most prominent for breakthrough patents in the 2000s and 2010s, accounting for 37 percent

[18]The Information sector includes Software Publishing (5112), Telecommunications (5173-79), and Data Processing (5182). The constituents of Professional Services are listed in footnote 15.

of breakthroughs in the last decade. We summarize these stark patterns in our first fact:

Fact 1. *US patenting was dominated by manufacturing firms from 1977 until 2000. During the 2000s, manufacturing firms' shares of patenting activities fell to 42 percent of total patents, 33 percent of external citations, 16 percent of breakthrough patents, and 18 percent of patent value. By the 2010s, firms in Information, Professional Services, Management, and Wholesale dominate US patenting.*

The rise in patenting among Wholesalers and Professional Services firms is consistent with the increasing prevalence of factoryless goods producers – firms that design goods and coordinate the production process, but outsource physical transformation tasks to other firms, increasingly in other countries (Bernard and Fort, 2015; Kamal, 2023; Fort, 2023). The rising importance of firms with Management as their primary sector is also suggestive of US multinationals' strong role in US innovation, since past work documents their disproportionate employment in that sector (Kamal et al., 2022). An interesting question for future work is to assess whether and how these patenting firms are directly involved in foreign manufacturing activities.

The middle row in Figure 4 depicts the declining dominance of mega-firms – those with more than 10 thousand US employees – in US patenting. In the late 1970s, these firms accounted for the majority of all four patent measures: 58 percent of patents and citations, 69 percent of breakthroughs, and 89 percent of patent value. By the 2010s, their share of patents, citations, and value had fallen over 15 percent, to 48, 49, and 72 percent, respectively. These firms' declining shares of patents contrasts with their rising shares of employment. As shown in Figure 5, mega-firms employ 25 percent of workers in the 1970s versus 28 percent in the 2010s.

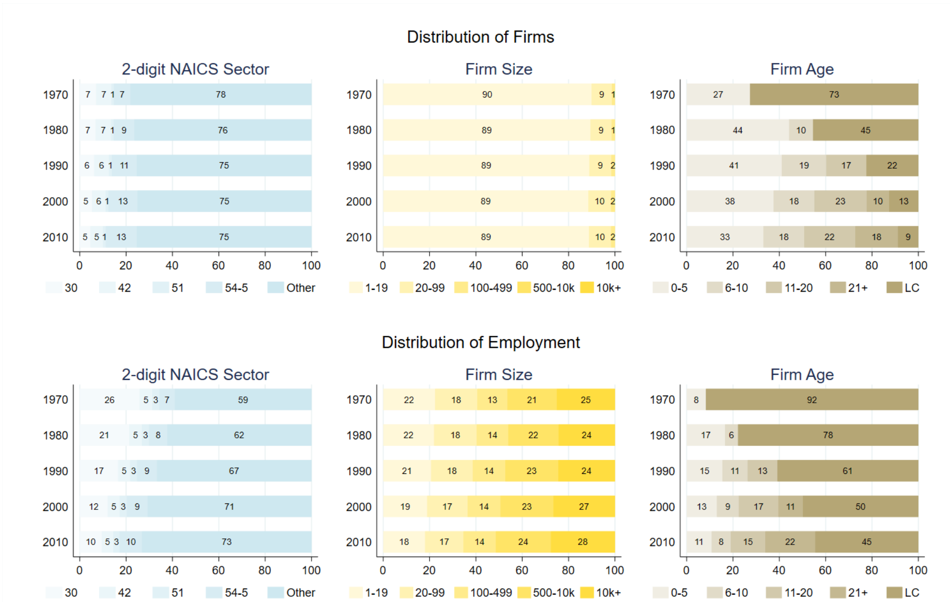
Mega-firms' share of breakthroughs evolves differently from other measures, as it is U-shaped over time: larger in the 1970s and 2000s, but smaller in the 1980s and 1990s, which aligns with evidence in Braguinsky et al. (2023).[19] Instead, firms with 100 to 500 employees almost double their share of 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. These changes likely reflect a combination of strong growth by 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.8).

Figure 4 also reveals a rising share of patent grants and citations by the *smallest* firms – those with fewer than 20 employees. Their share of patent grants increases steadily over the sample period, from 7 percent in the 1970s to 10 percent in the 2010s. This pattern is quite remarkable in light of the general shift towards large-old firms in the United States. For instance, as shown in Figure 5, the share of employment in such firms falls from 22 percent in 1977 to 18 percent over the period.

The bottom row of Figure 4 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 when

[19] Those authors define novel patents as those that combine two technology classes for the first time and find that mega-firms' importance in such patents also exhibits a u-shaped pattern over that period.

Figure 5: Firms and Employment by Firm Sector, Size and Age (BDS)



Source: BDS and authors' calculations. 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.

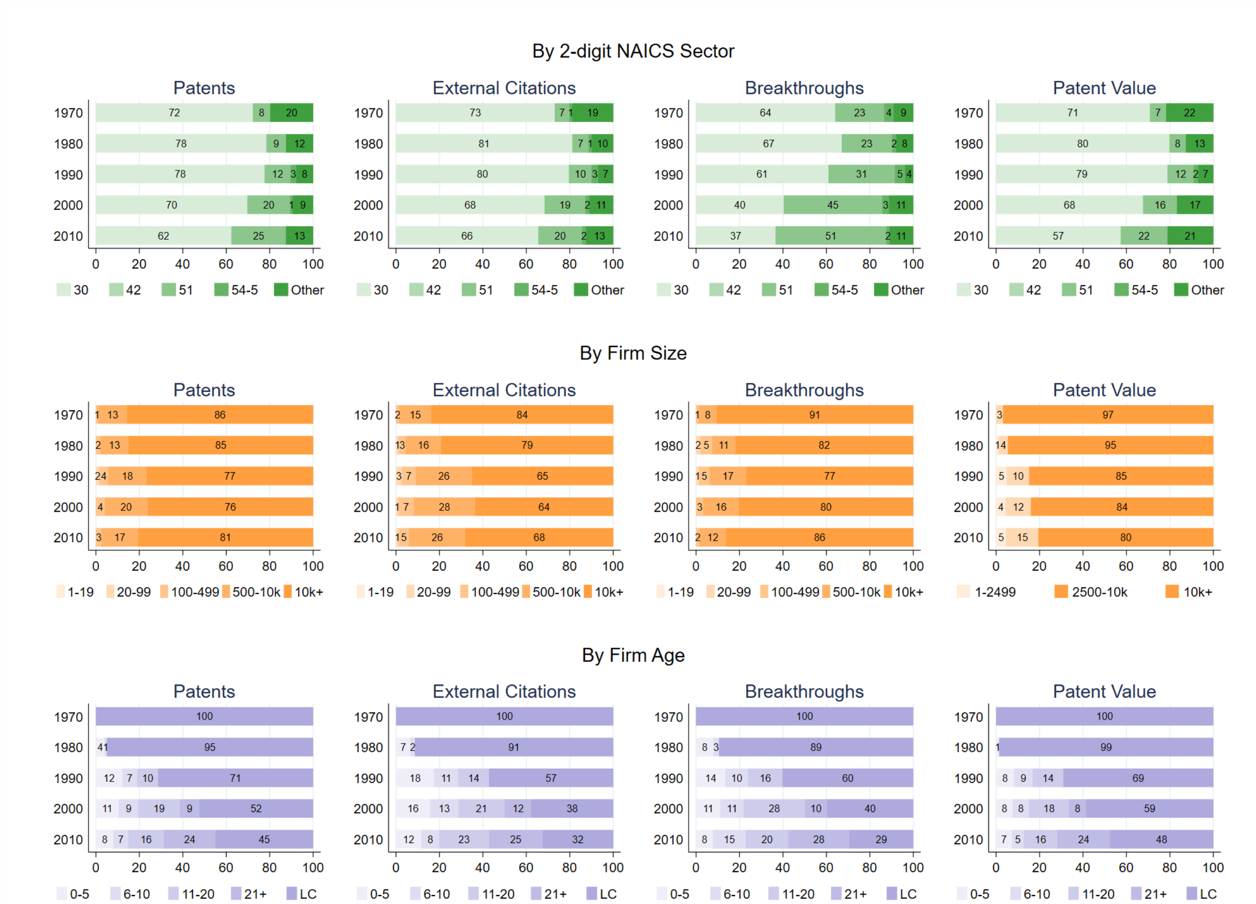
the firm's identifier (*firmid*) is first observed, age is left-censored in different years for different categories. By the 1990s, all firms can be categorized as at least 13 years old, so the first three bins are not censored.

There are several notable patterns about patenting and firm age. First, younger firms (0 to 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. Second, the oldest firms' increasingly produce a disproportionately low share of high-impact patents. By the 2010 decade, left-censored firms account for 55 percent of granted patents, versus just 49 and 48 percent of citations and breakthroughs. By contrast, firms between 11 to 20 years old account for 24 percent of breakthrough patents in the last decade, even though they receive just 13 percent of granted patents (and 15 percent of employment). The youngest firms (0 to 5 years) maintain a steady share of patent grants and citations from the 1980s to the 2010s (about 7 and 10 percent, respectively), while their share of breakthroughs drops from 9 to 5 percent. We summarize these findings on firm size and age in the following second fact:

Fact 2. *Large and old firms dominate US patenting, though mega-firms' (those with more than 10 thousand employees) share of patents and citations fall to less than 50 percent by the 2010s. Mega-firms are especially dominant in breakthrough patents, accounting for 69 percent in the late 1970s and 67 percent in the 2010s, though only 59 percent in the 1990s. Despite mega-firms' dominance of breakthroughs, medium-aged firms (11 to 19 years) are increasingly important for breakthrough patents, accounting for almost a quarter of breakthroughs compared to only 13 percent of total granted* *patents by the 2010 decade.*

The changing patterns of firm size and age in US patenting activity present a range of potential explanations for the changes in aggregate patents per R&D input documented in past work. Young and small firms may be better able to create new ideas, but less equipped to bring them to market such that they raise growth. If breakthroughs are the only patent types that lead to growth, then their level decline and lower mega-firm share in the 1990s could also be responsible for the apparent fall in researcher productivity in that period. The recovery by this small set of large firms may also signal a changing market structure in which innovators themselves grow, even as they acquire other firms and stifle competition. In the next section, we estimate the elasticity of patents to knowledge inputs in a manner that allows us to control for the types of firms that patent.

Figure 6: Patenting by Firm Sector, Size and Age (Compustat Panel)



Source: PV, KPSS, KPST, DS, Compustat and authors' calculations. Figure provides a breakdown of patents, external citations, breakthrough patents and patent value by Compustat firms' major 2-digit NAICS sector, size and age. Firms' sectors are time-invariant. Firm age represents time years Compustat. Bars representing less than 0.5 percent of patents are suppressed.

We conclude this section by illustrating the limitations of Compustat data for analyzing US patenting. Figure 6 replicates Figure 4, but using the Compustat panel. Three features of the Compustat data must be kept in mind when comparing the results. First, Compustat does not contain information on the full range of firms' establishments, so we rely on the single NAICS code reported for each firm in that dataset to determine its major sector.[20] Second, employment reported in Compustat can include workers outside the United States, whereas the LBD includes only US workers. Finally, firm entry in Compustat indicates going public (i.e., entering Compustat), not birth. As a result, age captures the firm's length of time as a public company, instead of its time in existence.[21]

With these caveats in mind, Figure 6 demonstrates that patenting activity among Compustat firms differs markedly from that of the LBD panel. In terms of sector and size, it tilts more substantially towards manufacturing and large firms. Indeed, in the final decade of the sample manufacturing firms and mega-firms account for 60 and 80 percent of patents in the Compustat panel versus 30 and 48 percent in the LBD panel. Young firms appear to patent more in the Compustat panel, though this pattern most likely reflects our inability to measure age prior to going to public in this sample. We summarize these findings in a third fact:

Fact 3. *Compustat data miss the massive shift in US patenting away from manufacturing firms over the last four decades: those data overstate manufacturing firms' share of US patent grants and citations by 30 and 41 percentage points in 2010. Compustat data also miss the substantial decline in mega-firms' (those with employment greater than 10 thousand) patenting shares, overstating their dominance in the 2010s by 28 percentage points for breakthroughs to 67 percentage points for granted patents.*[22] *Such firms account for 48, 49, and 67 percent of US patents, citations, and breakthroughs among the universe of US firms in 2010, versus 80, 79, and 86 percent among publicly traded firms.*

Having established two new facts about the evolution of firms involved in US patenting, and demonstrated the limitations of documenting such patterns using the subset of publicly traded firms, we now turn to analyzing whether and how firms' ability to translate knowledge inputs into patents has changed over this period.

5 Estimating Knowledge Elasticities

In this section, we analyze whether and how the efficiency of firm patenting with respect to R&D inputs has changed over time. We first provide simple "efficiency" figures that depict different measures of patents per real R&D expenditure by semi-decade. We next exploit the detailed firm-level data to estimate time-varying patent elasticities, controlling for both firm and time fixed effects. We then discuss how these elasticities and fixed effects are interpreted through the lens of the macro-growth models.

[20] We use the time-invariant, contemporaneous NAICS codes provided by Compustat (*naics*) in constructing this figure. While Compustat also reports historical NAICS codes (*naicsh*), these are missing for approximately one fifth of our panel. In a future iteration of this draft, we plan to concord the historic SIC codes (*sich*) to NAICS for this comparison, but note that using the contemporaneous NAICS should overstate Compustat firms' non-manufacturing activities in the past, but should not miss the shift out of manufacturing in later years.

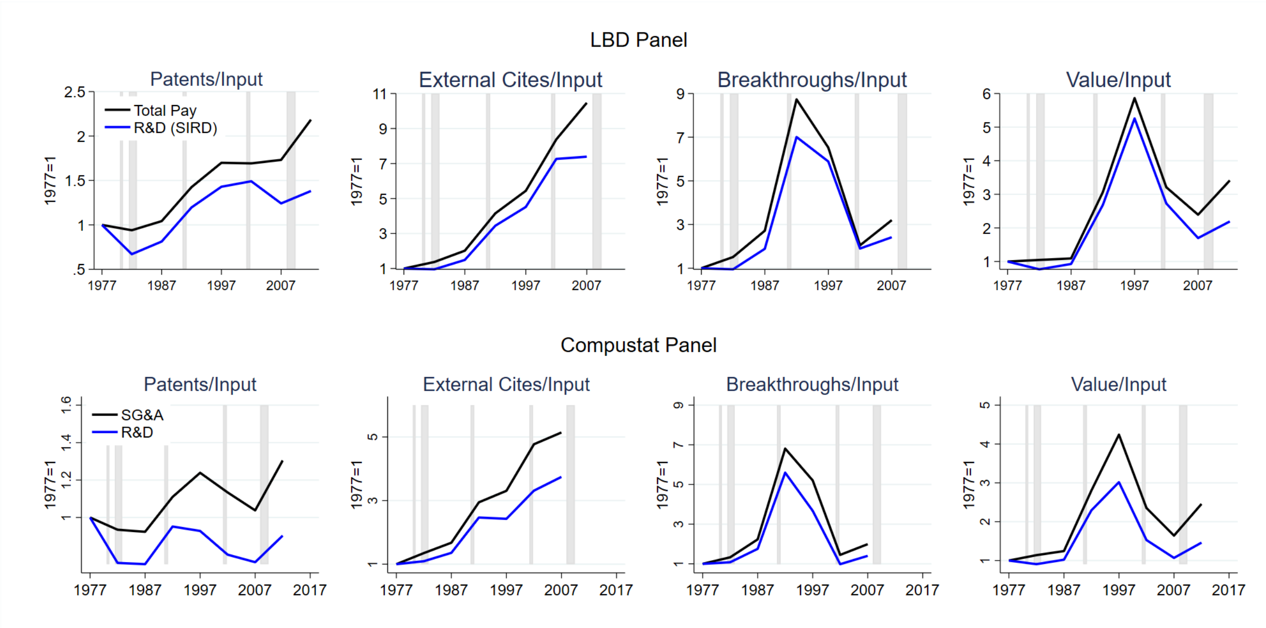
[21] While Loughran and Ritter (2004) provide data on Compustat firms' actual founding dates, this information is only available for about one third of the observations in our 1977 to 2017 sample.

[22] (80-48)/48=0.67 and (86-67)/67=0.28.

5.1 Patents per R&D Input

The top row of Figure 7 presents "simple" patent efficiency – patent activity per knowledge *flow* – over time in the LBD. We minimize Census disclosure burden, as above, by reporting results by semi-decade, and by displaying trends only with respect to the broadest and narrowest knowledge inputs: real total payroll and real R&D expenditure.[23] All firms are included in the former, while the latter is restricted to the firms in each year that appear in the RADS surveys. All ratios are indexed to 1 in the first, 1977 to 1981 semi-decade.

Figure 7: "Simple" Patent Efficiency



Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, DS, Compustat and authors' calculations. Notes: Top row reports patents, breakthroughs, external citations, and patent value per payroll or real R&D expenditures for the LBD panel. Second row reports analogous ratios with respect to real SG&A and R&D expenditures for 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. In the LBD panel, the set of firms included in the R&D line is restricted to those appearing in the R&D surveys in each year.

The first panel in Figure 7 depicts total granted patents per real R&D dollar. This measure falls from 1977 to 1982, grows steadily during the 1980s and 1990s, declines after 2002, and starts recovering after 2007.[24] Patents per total real payroll seem to grow more steadily throughout the period, with the exception of the 1997 to 2007 decade when they are flat. This divergence can be explained by: a) differences in the samples – the total payroll line includes all firms' US patents while

[23] 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.

[24] Recall that we present forward-looking five-year averages, so the 2002 number captures the average ratio between 2002 and 2007.

the blue line encompasses only patents by firms in the RADS surveys; and b) the relatively faster growth of R&D expenditure than payroll depicted in Figure 3.

This first panel in the bottom row of Figure 7 depicts analogous efficiency measures for the subset of Compustat firms, but with very different patterns. Patents per R&D fall initially, rise from 1987 to 1992, and then fall until 2007 after which they start to recover. Patents per R&D expenditure by 2012 are almost on par with the initial 1977 value. The pattern for SG&A is similar, though with more growth such that the final level is about 1.3 times the initial level. For patents per R&D expenditure, the overall decline between 1977 and 2007 aligns with past evidence used to motivate explanations for the deterioration of US researchers' ability to generate ideas (e.g., Kortum, 1997).

In contrast to the declining number of patents granted per R&D dollar, the second panel of Figure 7 shows that external citations per R&D dollar rise more or less steadily among both LBD and Compustat firms, though growth tails off between 1992 and 1997 among the Compustat firms and between 2002 and 2007 among LBD firms. There is no such decline for external cites per total payroll in the LBD or per SG&A in Compustat.

The last two panels in Figure 7 depict breakthrough patents and patent value per R&D expenditure. These panels paint a very different portrait of patent efficiency. The number of breakthrough patents per R&D dollar surges from 1987 to 1992 and then plummets back to about its 1987 level by 2002 for both LBD and Compustat firms. Patent value per R&D dollar displays a similar pattern, though the peak and rise occur five years later. These spikes align with the internet revolution and dot-com frenzy of the late 1990s, suggesting a link to the arrival of an entirely new set of technologies. The lagged response of patent value suggests it may take time for firms to figure out how to use new technologies or for investors to factor them into their growth expectations.

The depictions of patent activity per R&D expenditure in Figure 7 reinforce the fact that different measures and samples of R&D inputs and outputs can change the conclusions we draw on researcher efficiency. We therefore turn to estimating patent elasticities in a regression framework that allows us to incorporate firm and year fixed effects that might obscure changes in this fundamental relationship.

5.2 Patent Elasticity Estimates

We estimate time-varying elasticities of firm-level patents to R&D expenditures using Poisson-Pseudo Maximum Likelihood (PPML),

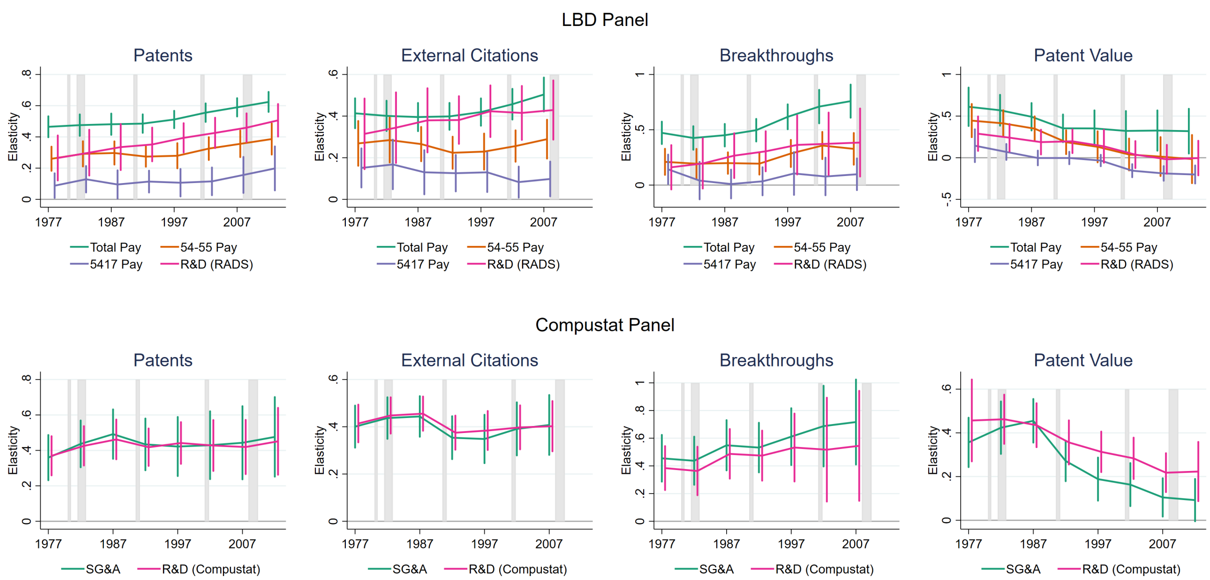
$$Y_{ft} = exp\left(\sum_{j=1977(5)k} \eta_j d[j]_{ft} \times ln(KnowledgeStock_{ft}) + \sum_{j=1977(5)k} \gamma_j d[j]_{ft} + \gamma_f + \epsilon_{ft} \right), \quad (6)$$

where Y_{ft} is firm f's patent output – patent grants, breakthroughs, external citations, or value – in year t. Dummies $d[j]_{ft}$ are equal to 1 for years t in semi-decade j. For example, $j = 1977$ encompasses years 1977 to 1981, while $j = 2012$ represents years 2012 to 2016. The sample period is 1977 to 2016 for patent grants and patent value, and 1977 to 2011 for the five-year forward-looking breakthroughs and external citations. The last semi-decade in the sample is therefore either k=2007 or k=2012. η_j

is the elasticity of interest, capturing the marginal change in patent output with respect to a change in firm f's real knowledge-input stock during semi-decade j. γ_j are semi-decade fixed effects that capture average idea creation by semi-decade that is common to all firms in the sample in those years.[25] γ_f capture average patents over the entire sample period for each firm in the sample.

We consider four real knowledge-input stocks in the LBD panel: capitalized total payroll, Professional Services and Management (NAICS 54-55) payroll, R&D lab (NAICS 5417) payroll, and R&D expenditure from RADS.[26] In each case, the set of firms included in the estimation is restricted to those for which we observe an input *flow* in that year, as we set stocks to missing (as discussed in Section 3.6) 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. Standard errors are clustered by firm.

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 6 by patenting activity and knowledge stock for the LBD and Compustat panels. Whiskers denote 95 percent confidence intervals. Standard errors are clustered at the firm-level.

Surprisingly, we find flat or *rising* elasticities for patents, external citations, and breakthroughs in almost all cases. The top row of Figure 8 plots our estimates of η_j for the LBD panel. The first panel in this row depicts a fairly steady increase in firms' patent elasticities over time across all four knowledge stocks. For total pay, the elasticity rises from 0.46 to 0.62 between the first and last

[25]We estimate elasticities at the semi-decade frequency and use semi-decade fixed effects to minimize Census disclosure burden. Estimations employing year fixed effects yield analogous conclusions.

[26]We use knowledge input stocks rather than flows to capture the idea that flows in one year may affect idea creation for several years. Hausman et al. (1984) explore different lag structures and find that such lags do not add much explanatory power once firm fixed effects are included.

semi-decade, indicating that a 1 log point increase in payroll is associated with 0.16 (=0.62-0.46) log point more patents at the end versus beginning of our sample period.

The sizes of our estimated elasticities for R&D expenditures here are in line with those based on R&D *flows* in Hausman et al. (1984), who regress patents for 128 firms from 1968 to 1974 on contemporaneous R&D expenditures 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 sample period. We estimate an elasticity of 0.26 for R&D (RADS) expenditures in the 1977 to 1982 period, versus 0.48 in the 2010 decade. The magnitudes of our coefficient estimates are thus similar, even as their evolution moves in the opposite direction in our later and much longer sample period.

The trends in elasticities for patents weighted by external citations in the second panel are flatter than for patenting, though still flat or rising for all inputs except R&D lab (NAICS 5417) pay. The elasticities for breakthrough patents in the third panel are flat for that knowledge stock, while rising for the other three. Across the four knowledge stocks, patent elasticities are generally highest for the total payroll stock and lowest for R&D lab (NAICS 5417) pay. These differences likely arise because of the very different levels in patenting across firms. The regression sample for the total pay elasticities includes all firms in the US economy, the vast majority of which do not patent. By contrast, the sample for the R&D lab (NAICS 5417) elasticities is restricted to firms with R&D lab (NAICS 5417) employment, which is a small subset of firms that perform R&D. Mean patenting among all firms is clearly lower than among firms with an R&D lab (NAICS 5417).

The first three panels in the first row of Figure 8 suggest that among LBD firms, R&D inputs are at least as efficient at producing ideas by the 2000s as they were in the 1970s. The corresponding panels in the second row of Figure 8 indicate that this trend is also true among Compustat firms. This similarity in LBD and Compustat regression results here – in contrast to the descriptive statistics discussed in Section 4 and the simple patenting per input flows displayed in Figure 7 – highlights the importance of using firm and year fixed effects to better understand the way in which ideas might be getting harder to find. Indeed, for assessing researchers' efficiency, we get a remarkably similar message across knowledge inputs and firm samples.

As a first step towards assessing the link between knowledge inputs and *growth*, the final panel of each row of Figure 8 displays analogous estimated elasticities for patent value. Recall that patent value captures both the granting of a patent *and* investors' expectation of the future value of that patent to the patenting firm. In contrast to the flat or rising trends found for patents, external citations, and breakthroughs, these elasticities decline noticeably over time. For the total payroll stock, they fall by half, from 0.61 to 0.32, while for the Professional Services and Management (NAICS 54-55) payroll and R&D expenditures stocks, they decline from 0.43 and 0.28, respectively, in the initial period to zero. For the R&D lab (NAICS 5417) payroll stock, they fall below zero to -0.19 in the final period, suggesting additional R&D lab (NAICS 5417) workers lead to *less* valuable patents. Results are comparable among Compustat firms.[27]

The starkly different trends for patent value versus patents, external citations, and breakthroughs suggests that firms may find it increasingly difficult to translate ideas into growth, even as researchers' efficiency improves.[28] This difficulty may be driven by several factors. New ideas may cannibalize old ones, both within and across firms (e.g., Antràs et al., 2024). Evidence in favor of this mechanism is presented by KPSS, who show that firm growth responds negatively to greater patent value creation among competitor firms in the same sector. As technologies mature, such cannibalization effects may get stronger and dominate any potential complementarities. Another possibility is that new ideas lead to technological change, which in turn alters the innovators' competitive environment. For example, the internet revolution ushered in a new technology sector in which production features high fixed costs, low marginal costs, and significant network effects. That sector is now dominated by handful of large firms with considerable market power and resources, which may provide incentives and capabilities to restrain output and stifle competition, for instance via acquisitions (e.g., Cabral, 2024). In Section 6, we estimate the relationship between ideas and growth directly. Before that, however, we discuss the economic interpretation of our patent elasticity estimates.

5.3 Mapping the Estimates back to Theory

The assertion that "ideas are getting to harder to find" is often interpreted as a decline in researcher productivity, which might be interpreted as a decline in Romer's δ in equation 1.[29] In contrast to this interpretation, the positive and generally increasing patent elasticity estimates ($\hat{\eta}_j$) displayed in the first three panels of Figure 8 point to flat or rising researcher efficiency.

An alternative interpretation of the claim that "ideas are getting harder to find" is that the average number of ideas per firm is falling, even after controlling for firms' average patenting abilities and changes in their use of knowledge inputs. Under this interpretation, "ideas getting harder to find" implies a secular decline in the time fixed effects from equation (6) that affects all firms equally, regardless of their knowledge-input use. In the semi-endogenous growth models, decreases in these time effects are driven solely by the rising stock of existing ideas (i.e., $A^{-\beta}$, where β measures the concavity in the growth rate of new ideas as a function of the existing stock). In practice, the estimated time effects also capture changes in patenting laws, macro-economic fluctuations, or any other shocks that affect all firms similarly.

To make further progress on understanding whether "ideas might be getting harder to find," we broaden our discussion of the regression results from the last section to consider how the patent elasticity estimates, the time fixed effects, and the firm fixed effects combine to predict actual patenting. The firm fixed effects capture average patenting over the period by firms in the sample, after controlling for variation in their use of knowledge inputs. While these effects are time-invariant for each

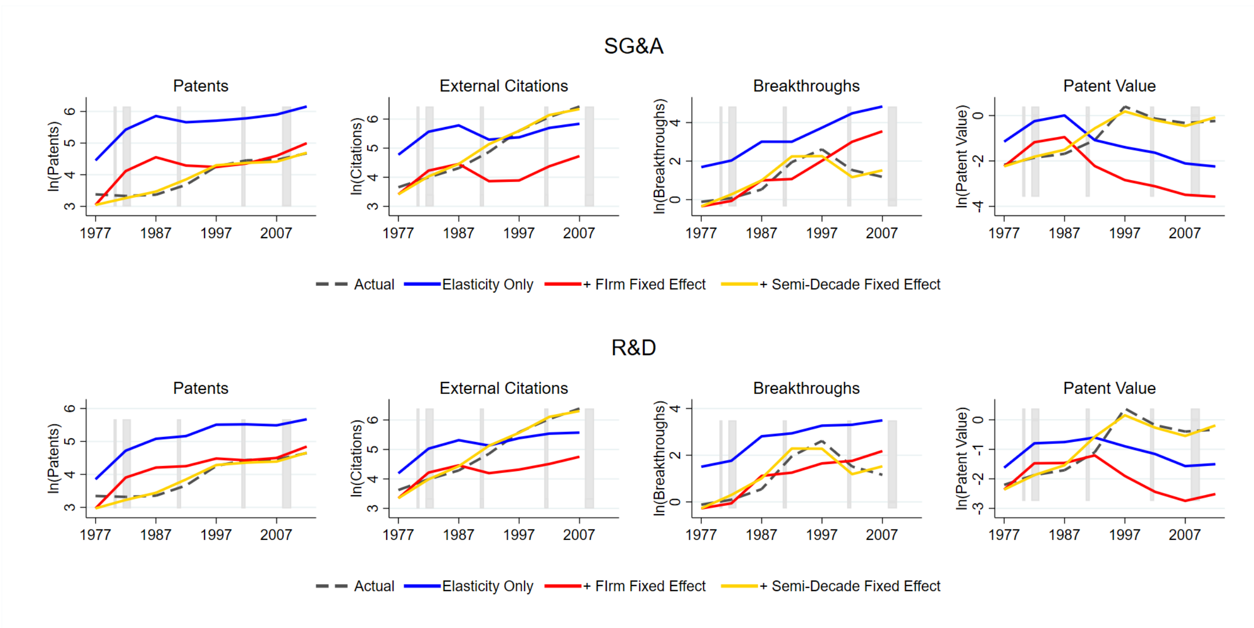
[27] In both panels, we obtain similar results for patents, external citations, and breakthroughs when restricting estimation for those outcomes to the set of patents and firms for which we observe value. For the LBD panel these results are not yet disclosed. For the Computsat panel, see Appendix Figure A.9.

[28] We investigate the relationship between knowledge inputs and patent outcomes further by estimating separate elasticities for *counts* of low- versus high-citation, -novelty, and -value patents in Appendix H.

[29] For example, Bloom et al. (2020) state in their abstract "...researcher efforts are rising substantially while researcher productivity is declining sharply..."

firm, their contribution to predicted aggregate patenting will vary over time as the sample composition changes. The time fixed effects measure average patents in each semi-decade by all the firms in a particular sample, again after controlling for changes in their use of knowledge inputs as well as average firm patenting via the firm fixed effects. These semi-decade fixed effects thus capture variation over time in patenting common to all firms, as well as any compositional differences in patenting growth rates by entrants and exiters for a particular sample.

Figure 9: Predicted Growth in Patenting from Elasticity Estimates (Compustat Panel)



Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, Compustat and authors' calculations. Notes: Figure reports growth in patents, external citations, breakthrough patents and patent value over the 1977 to 2007 (external citations, breakthroughs) or 2012 (patents, patent value) sample period.

Figure 9 displays the role of each of these factors in matching the data by plotting actual patents by Compustat firms, along with their predicted patents as each of the factors above is added sequentially.[30] The blue lines plot predicted patents using firms' observed knowledge inputs and estimated patent elasticities. For both patent grants and breakthroughs, the rising elasticities and knowledge-input use over-predict firms' patenting output. Adding in the firm fixed effects generally shifts the predicted line down to match the level of patenting by Compustat firms, without altering its shape significantly (red line). Finally, adding the semi-decade fixed effects changes the overall trend so that it aligns with the actual data (yellow line).

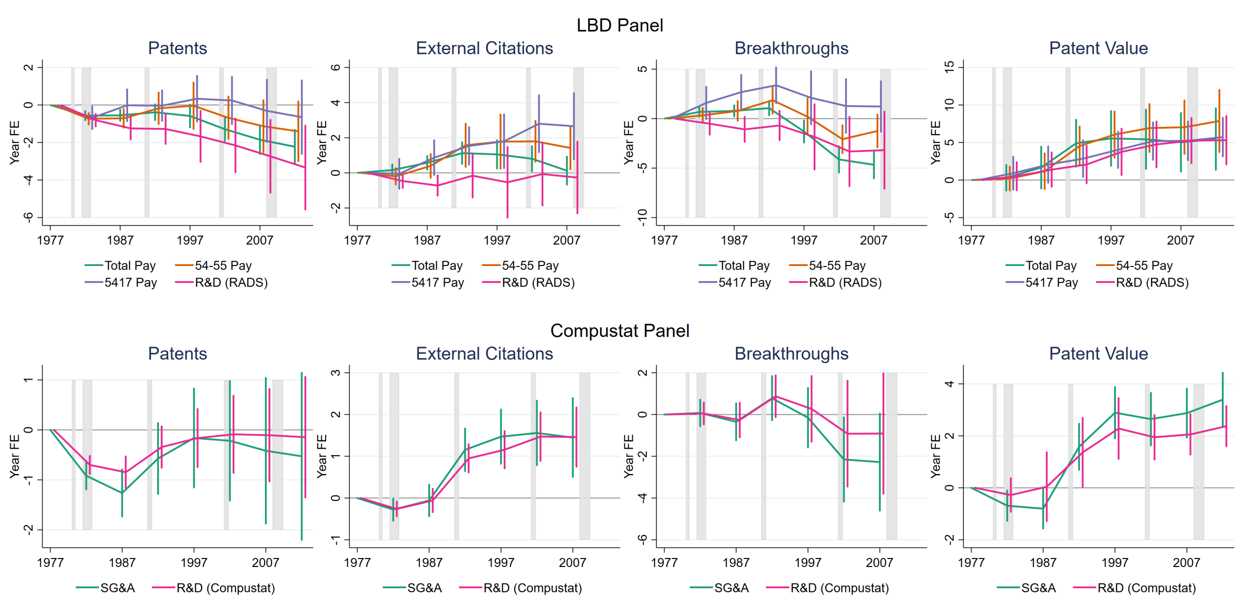
We emphasize two main points from this Figure. First, the semi-decade effects play a key role in reconciling the predicted time-series behavior of patenting activity with what is actually observed in the data. As explained above, steady decreases over time and across samples in these time fixed

[30]We present the results for Compustat firms only at this stage to minimize disclosure. If these figures prove compelling in presentations, we will also construct them using the LBD panel.

effects could reflect the type of concavity in the production of ideas suggested by Jones (1995). Firms might still patent relatively more in those periods in which they use more knowledge inputs, even as overall patenting for all firms, regardless of their innovative efforts, declines. If the time fixed effects fall more as the aggregate stock of ideas grows, this would support the notion of existing ideas 'crowding out' new ideas captured by $\beta > 0$ in equation (3). The time fixed effects could also vary over the period due to changes in patenting laws, changes in firms' desire to protect innovations via patents, or even entry or exit into the sample of firms with different patenting growth rates.

To examine these time effects more closely, we plot them in Figure 10 for both the LBD and Compustat samples. The first panel in the top row displays steadily falling semi-decade fixed effects for patent grants by all firms (total pay line) and firms in the RADS (R&D line). By contrast, there is no statistically significant decline in the time fixed effects for firms with Professional and Scientific (NAICS 54-55) pay or R&D lab (NAICS 5417) pay. Compustat firms' semi-decade fixed effects initially decline but then rise back to their initial level, consistent with the over-prediction of patenting based solely on estimated elasticities (blue line) in Figure 9.

Figure 10: Estimated Semi-Decade Fixed Effects (γ_j)



Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, Compustat and authors' calculations. Notes: Figure reports estimates of γ_j from Equation 6 by patenting activity and knowledge stock for the LBD and Compustat panels. Whiskers denote 95 percent confidence intervals. Standard errors are clustered at the firm-level.

The second panel of Figure 10 displays flat or rising time effects for external citations, with particularly strong growth among Compustat firms. Recall that this measure contains external citations over a five-year window after the patent's grant, so it will not automatically rise over time. Indeed, Appendix Figure A.14 plots the same estimates using a balanced panel of Compustat firms and displays relatively flat semi-decade fixed-effect estimates, suggesting that the growth here arises from the changing composition of the sample. On average, Compustat firms are accruing grants with higher citation counts over the period, even after controlling for their use of knowledge inputs and the average patenting of firms in the sample. The semi-decade fixed effects for patent value display even more consistent increases over time. The stock market's expected returns from new patent grants thus exhibit a secular rise for all publicly traded firms, even as the expected returns from firms' R&D efforts decline (Figure 8).

The second key message from Figure 9 is that all firms experience a surge and subsequent decline in breakthrough patents midway through our sample period that is not predicted by their knowledge-input use. This surge – also evident in the semi-decade fixed effect estimates in Figure 10 – occurs during the internet boom, suggesting the possibility that a new class of ideas demonstrates initially explosive growth and then tails off over time. Such a tailing-off seems precisely in line with the case study evidence in Bloom et al. (2020). Crucially, however, this tailing-off need not continue indefinitely, since the discovery of an entirely new area remains possible (e.g., the recent AI boom). Indeed, Evenson and Kislev (1976) model applied research as having diminishing returns within the distribution of a particular set of ideas, while basic research is used to discover new distributions. We note that the semi-decade fixed effects recover to their initial levels for some of the firm samples by 2007 (e.g., NAICS 54 -55 and 5417 pay).

To summarize, US researchers' patent efficiency ($\hat{\eta}_j$) – one of the ways of interpreting "ideas are getting harder to find" – does not decline over our 46-year sample period. If anything, the elasticity of ideas to R&D efforts has risen. This fact, however, is perhaps not what the semi-endogenous growth models have in mind. Indeed, those models assume that Romer's δ (i.e., researcher productivity) is constant over time and instead posit a declining growth rate of ideas in response to a growing stock of ideas. While the semi-decade fixed effects ($\hat{\gamma}_j$) for patent grants and breakthroughs decrease for some specifications after 1997, this pattern is not universally true across all measures of knowledge inputs. Moreover, citations and patent value exhibit flat or *rising* secular trends. These patterns are inconsistent with a secular decline in patenting activity as the aggregate stock of ideas grows, motivating our examination of the link between ideas and growth in the next section.

6 Estimating the Link Between Growth and Ideas

The previous section demonstrates that researcher efficiency in terms of the elasticity of patenting to knowledge input stocks ($\hat{\eta}_j$) is flat or rising over the last five decades. It also shows that average patenting by semi-decade across firms, after controlling for knowledge-input use and firm fixed effects, declines for some measures of patenting but is flat or rising for others. Given these results, why has aggregate growth has not kept pace with rising researcher efforts?

To address this question, we now examine the link between ideas and growth. A key assumption in many of macro-growth models is that growth in ideas ($\dot{A}$) maps one-to-one with growth in output, and that this relationship is constant over time.[31] Yet as noted in the introduction, the evidence connecting ideas to growth is fairly limited. A better understanding of this relationship may be

[31] For example, Bloom et al. (2020) note that they "...follow much of the literature...and define ideas to be in units so that a constant flow of new ideas leads to constant exponential growth in A" (p. 1108).

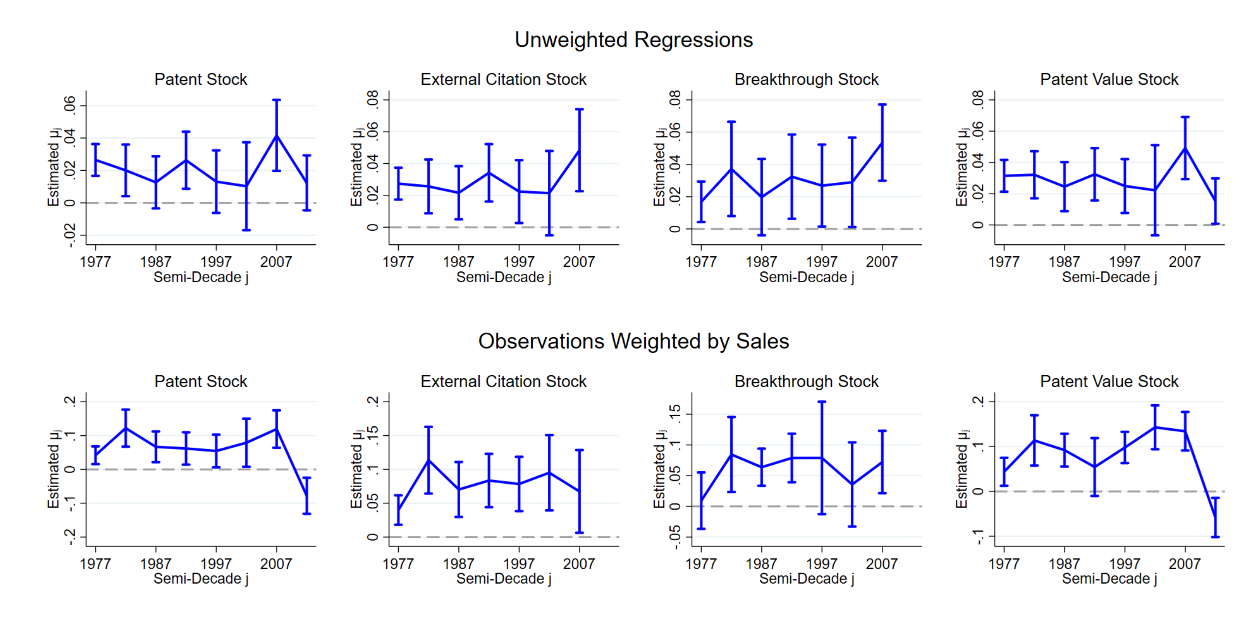
helpful in inspiring additional models of long-run growth.

Toward that end, we examine the link between patenting and firm growth, allowing the relationship to vary over time. Specifically, we estimate:

$$DHS\left(X_f^{t:t+\tau}\right) = \alpha DHS\left(Employment_f^{t:t+\tau}\right) + \sum_{j=1977(5)k} \mu_j^\tau d[j]_{f,t} \times DHS\left(IdeaStock_f^{t:t+\tau}\right) + \sum_{j=1977(5)k} \phi_j^\tau d[j]_{ft} + \sum_{b=1:5} \rho_k^\tau 1\{SizeBin_b\} + \nu_{ft}^\tau, \tag{7}$$

where the left hand side is the Davis et al. (1996) (DHS) growth rate of firm f's sales between years t and $t+\tau$. The first term on the right-hand-side of the equation is the contemporaneous DHS growth rate in employment. μ_j^τ is the elasticity of interest, capturing the relationship between sales growth and firm's contemporaneous growth in their stock of ideas.[32] As in our earlier regressions, ϕ_j^τ represent semi-decade fixed effects. These coefficients capture time-varying changes in growth common to all firms. The final term on the right-hand-side is a control for firm size.

Figure 11: Time-Varying Estimates of μ_j^τ from Equation 7



Source: PV, KPSS, KPST, Compustat and authors' calculations. Notes: Figure reports relationship between Compustat firms sales and idea stock growth between years t and $t+4$ (μ^4 from Equation 7). Whiskers denote 95 percent confidence intervals. Standard errors are clustered at the firm-level.

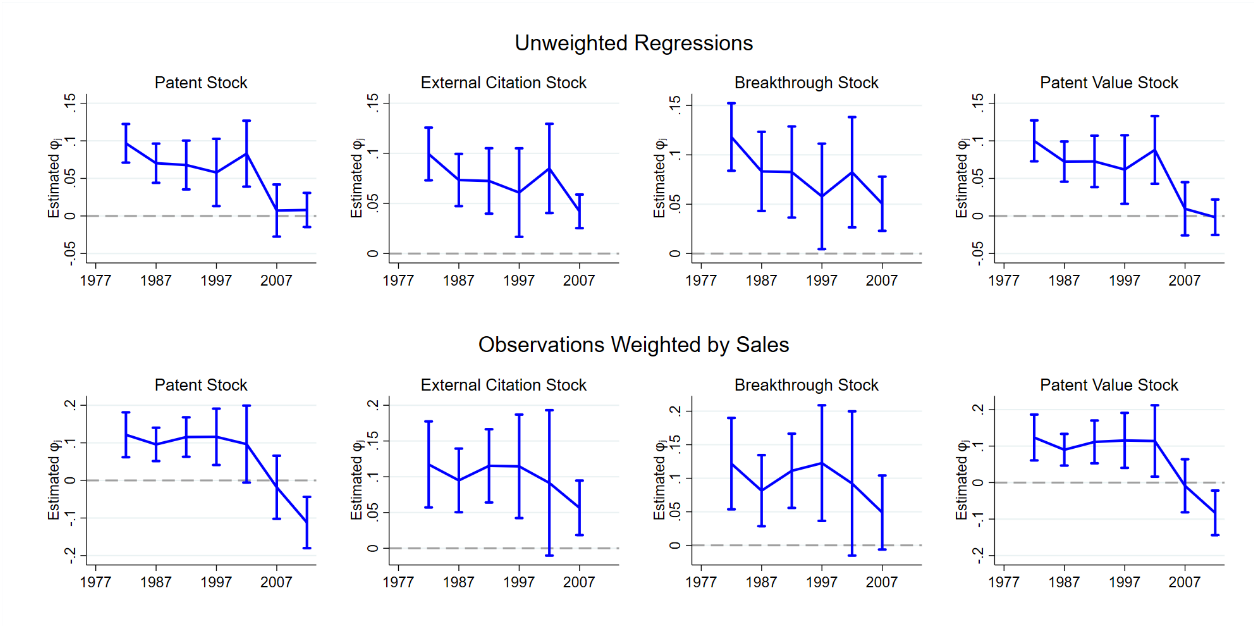
At present, we include results only for our Compustat panel. We plan to add analogous estimations using the LBD in a future draft. Estimates of the relationship between sales and ideas for $\tau=4$ (μ_j^4) along with their 95 percent confidence intervals, are reported in Figure 11. As indicated in the figure, we generally find a positive relationship between the two variables across idea stock measures over

[32] Idea stocks are computed analogously to the knowledge stocks considered in the last section.

our sample period, with no real trend.

Figure 12 reports the semi-decade fixed effects from estimation of equation 7, along with their 95 percent confidence intervals. As indicated in the figure, we generally find a *negative* trend in these fixed effects over time, indicating that growth, conditional on ideas, is getting harder to find.

Figure 12: Year Fixed Effects from Equation 7



Source: PV, KPSS, KPST, Compustat and authors' calculations. Notes: Figure reports the semi-decade fixed effects for τ=4 (ϕ_j^4 from Equation 7). Whiskers denote 95 percent confidence intervals. Standard errors are clustered at the firm-level.

7 Conclusion

We build a novel 46-year firm-level panel to study the evolution of US patenting. In contrast to the notion that ideas are getting harder to find, we show that the elasticities of patent grants, external citations, and breakthrough patents with respect to six different measures of R&D inputs are flat or rising over time. Researcher efficiency has not declined over the last five decades.

We also estimate semi-decade fixed for these measures of US patenting activity across five samples of firms and for the same six knowledge inputs. These fixed effects capture average patenting in a particular sample and time period, after controlling for those firms' average patenting and changes in knowledge inputs. A secular decline in these time effects might reflect an aggregate 'crowding out' of ideas as the pool of total ideas grows. While the estimated semi-decade fixed effects decline steadily for patent grants and breakthrough patents in two samples of firms, they do not for others. Most notably, the estimated time effects for external citations and patent value are flat or rising on average over time.

To reconcile these patterns with the fact that aggregate growth rates have held fairly steady even as research inputs have increased, we study the mapping between ideas to growth. In line with the observation in Romer (1990) that ideas may be complements, substitutes, or a mix of both in final-good production, we document considerable variation in the relationship between ideas and growth, both across different types of ideas and over time. These results point to a new path for macro-growth models to explore when reconciling their aggregate predictions to the data.

Our analysis provides an optimistic view of human's innovative capacity and our ability to sustain long-term growth. The semi-endogenous growth models assume that ideas get harder and harder to find as our stock of knowledge grows and imply that population growth is the only way humans can continue to improve their living standards. By contrast, our analysis shows that, at least over the last five decades, ideas don't seem to crowd each other out.

References

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. Econometrica 60(2), 323–351.

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

Antràs, P., E. Fadeev, T. C. Fort, and F. Tintelnot (2024). Export-platform fdi: Cannibalization or complementarity. AEA Papers and Proceedings 114(4), 130–135.

Balasubramanian, N. and J. Sivadasan (2011, May). What happens when firms patent? new evidence from us economic census data. Review of Economics and Statistics 93(1), 126–146.

Bernard, A. B. and T. C. Fort (2015, May). Factoryless goods producing firms. American Economic Review 105(5), 518–23.

Bessen, J. (2008). The value of us patents by owner and patent characteristics. Research Policy 37(5), 932–945.

Bilir, L. K. and E. Morales (2020). Innovation in the global firm. Journal of Political Economy 128(4), 1566–1625.

Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2020, April). Are ideas getting harder to find? American Economic Review 110(4), 1104–44.

Braguinsky, S., J. Choi, Y. Ding, K. Jo, and S. Kim (2023). Mega firms and recent trends in us innovation: Empirical evidence from the us patent data. Technical Report 31460, NBER.

Brynjolfsson, E. and A. McAfee (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & company.

Cabral, L. M. B. (2024). Big tech acquisitions. International Journal of Industrial Organization.

Chow, M. C., T. C. Fort, C. Goetz, N. Goldschlag, J. Lawrence, E. R. Perlman, M. Stinson, and T. K. White (2021). Redesigning the longitudinal business database. Technical Report 28839, National Bureau of Economic Research Working Paper Nr 28839.

Cohen, W. M., R. Nelson, and J. P. Walsh (2000). Protecting their intellectual assets: Appropriability conditions and why us manufacturing firms patent (or not). (7552).

Corrado, C., C. Hulten, and D. Sichel (2005). Measuring capital and technology: an expanded framework. In Measuring capital in the new economy, pp. 11–46. University of Chicago Press.

Cowen, T. (2011). 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.

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

Dinlersoz, E., N. Goldschlag, A. Myers, and N. Zolas (2018). An anatomy of us firms seeking trademark registration. In Measuring and Accounting for Innovation in the 21st Century. University of Chicago Press.

Dreisigmeyer, D., N. Goldschlag, M. Krylova, W. Ouyang, and E. Perlman (2018). Building a better bridge: Improving patent assignee-firm links. Technical report, Center for Economic Studies CES-TN-2018-01.

Dyevre, A. and O. Seager (2024). Matching patents to publicly listed firms in the us: 1950-2020. Technical report, LSE.

Eisfeldt, A. L. and D. Papanikolaou (2014). The value and ownership of intangible capital. The American Economic Review 104(5), 189–194.

Ekerdt, L. K. (2024). The role of r&d factors in economic growth. CES Working paper 24-69, Center for Economic Studies.

Ekerdt, L. K. and K.-J. Wu (2024). Self-selection and the diminishing returns to research. Working paper, Center for Economic Studies.

Evenson, R. E. and Y. Kislev (1976). A stochastic model of applied research. Journal of Political Economy 84(2), 265–282.

Fadeev, E. (2023). Creative construction: Knowledge sharing and cooperation between firms. Technical report, Working paper, Duke University Fuqua School of Business.

Fort, T. C. (2023). The changing firm and country boundaries of us manufacturers in global value chains. Journal of Economic Perspectives 37(3), 31–58.

Fort, T. C. and S. D. Klimek (2018). The effects of industry classification changes on us employment composition. CES Workign Paper (18-23).

Galasso, A. and T. S. Simcoe (2011). Ceo overconfidence and innovation. Management science 57(8), 1469–1484.

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

Graham, S. J., C. Grim, T. Islam, A. C. Marco, and J. Miranda (2018). Business dynamics of innovating firms: Linking us patents with administrative data on workers and firms. Journal of Economics & Management Strategy 27(3), 372–402.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. The Bell Journal of Economics, 92–116.

Griliches, Z. (1998). Patent statistics as economic indicators: a survey. In R&D and productivity: the econometric evidence, pp. 287–343. University of Chicago Press.

Grossman, G. M. and E. Helpman (1991). Quality laddres in the theory of growth. Review of Economic Studies 58, 43–61.

Growiec, J., P. McAdam, and J. Mućk (2022). Are ideas really getting harder to find? r&d capital and the idea production function.

Hall, B. H., J. Mairesse, and P. Mohnen (1998). Measuring the returns to r&d. In Handbook in Economics, Volume 2, Chapter 24, pp. 1034–1082. Elsevier B.V.

Hausman, J., B. H. Hall, and Z. Griliches (1984). Econometric models for count data with an application to the patents-r&d relationship. Econometrica 52(4), 909–938.

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

Jarmin, R. S. and J. Miranda (2002). The longitudinal business database. CES Working Paper CES-WP-02-17.

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

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

Kamal, F. (2023). A portrait of us factoryless goods producers. In N. Ahmad, B. Moulton, J. D. Richardson, and P. V. D. Ven (Eds.), Measuring the Global Economy. University of Chicago Press.

Kamal, F., J. McCloskey, and W. Ouyang (2022). Multinational firms in the u.s. economy: Insights from newly integrated microdata. Working Paper 22-39, Center for Economic Studies.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021, September). Measuring technological innovation over the long run. American Economic Review: Insights 3(3), 303–20.

Kerr, W. R. and S. Fu (2008). The survey of industrial r&d—patent database link project. The Journal of Technology Transfer 33, 173–186.

Kline, P., N. Petkova, H. Williams, and O. Zidar (2019). Who Profits from Patents? Rent-Sharing at Innovative Firms. The Quarterly Journal of Economics 132(2), 1343.1404.

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

Kortum, S. S. (1997). Research, patenting, and technological change. Econometrica 65(6), 1389–1419.

Kuhn, J., K. Younge, and A. Marco (2020). Patent citations reexamined. The RAND Journal of Economics 51(1), 109–132.

Loughran, T. and J. Ritter (2004). Why has ipo underpricing changed over time? Financial management, 5–37.

Meyers, K. R. and L. Lanahan (2022). Estimating spillovers from publicly funded r&d: Evidence from the us department of energy. American Economic Review 112(7), 2393–2423.

Peters, B., M. J. Roberts, V. A. Vuong, and H. Fryges (2017). Estimating dynamic r&d choice: An analysis of costs and long-run benefits. RAND Journal of Economics 48(2), 400–437.

Ribeiro, B. (2024). Embracing the future or building on the past? growth with new and old technologies. Growth with New and Old Technologies (February 10, 2024).

Romer, P. M. (1990). Endogenous technological change. Journal of Political Economy 98(5), S71–S102.

Silva, J. S. and S. Tenreyro (2006). The log of gravity. Review of Economics and Statistics 88(4), 641–658.

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.

§†Yale 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.

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

[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.

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.
3. Select among multiple matches based upon string comparators.

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

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 firmids 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 firmids 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.

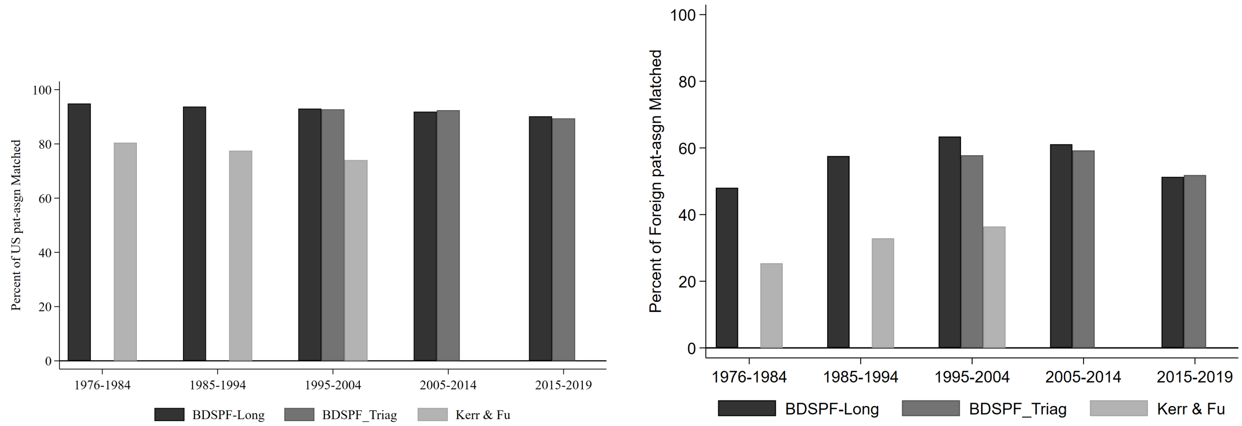
Table A.2: Raw Match Block Distribution

	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.

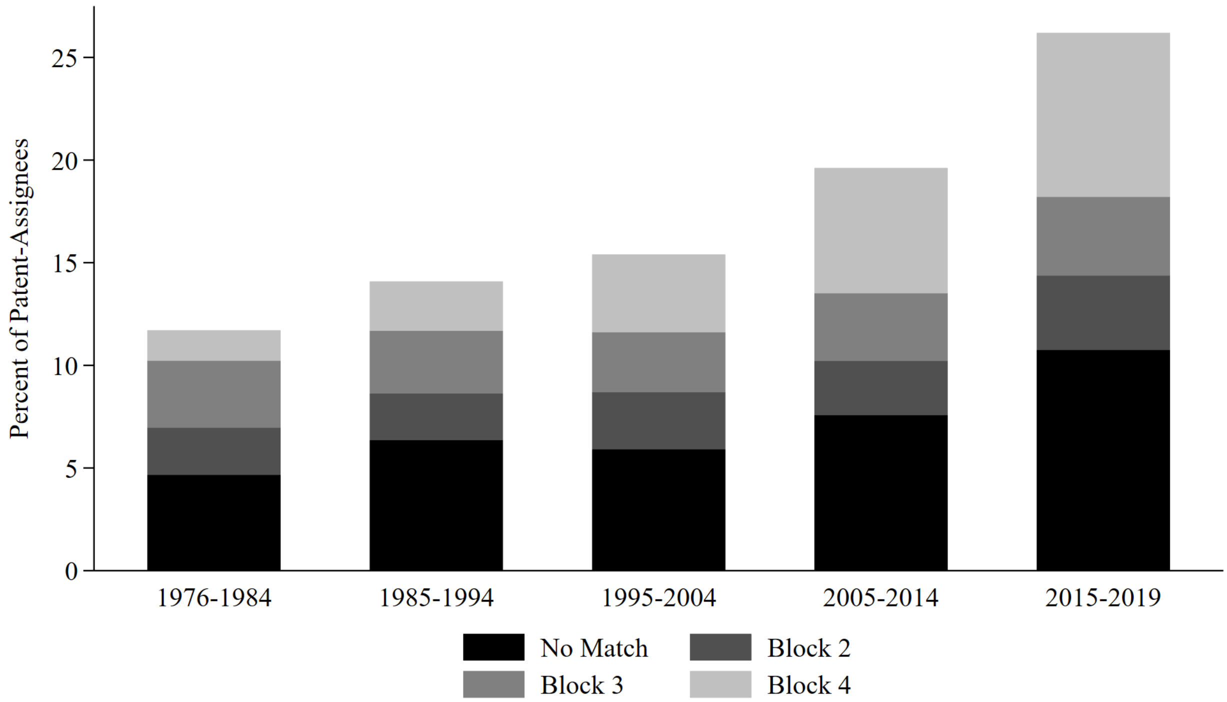
Table A.6: BDSPF-Long Match Precision & Recall

	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



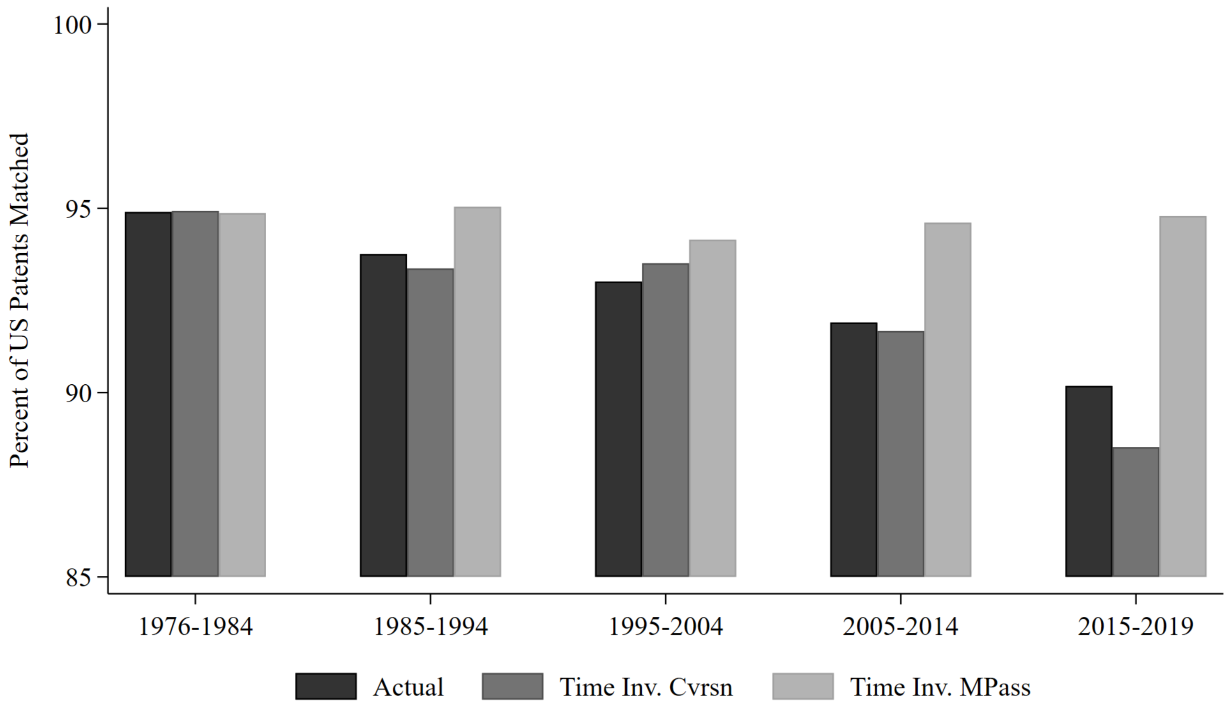
Source: BDSPF-Long, PatentsView
Notes: Figure shows the distribution of patent-assignee records (`patnum`, `assg_seq`) by the type of raw match, if any. The excluded category, which is also the largest, are those with a Block 1 match. If the Block 1 percent were included all groups would sum to 100 within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

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

[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.

(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 `firmid`s in either the `firmid` or `firmid_rorg` fields. `firmid`s 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_yr`s (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.

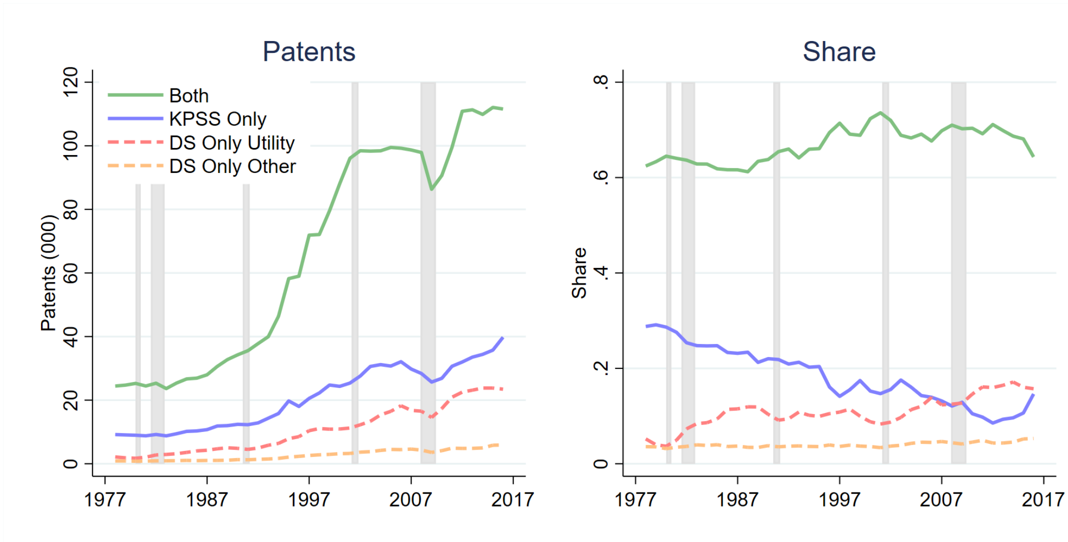
The stacked-line scatter plot in the left panel of Figure A.4 reports the cumulative number of patents matched to Compustat by either KPSS or DS in each application year, with the top line showing the total number of patents in the union of the two mappings, referred to as the "total" number of patents 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.

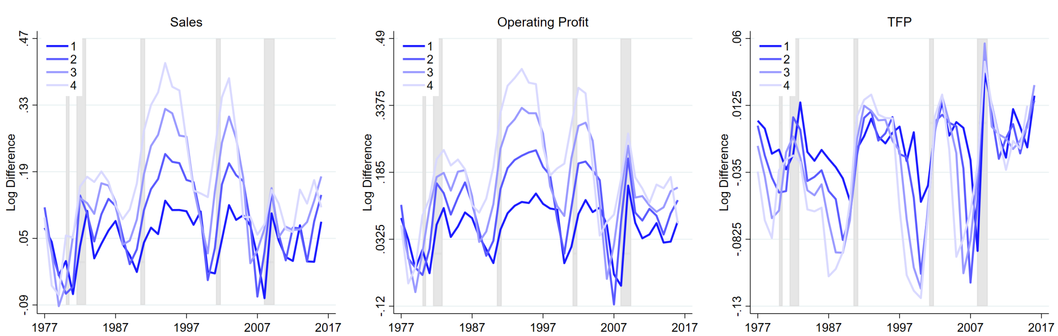
Figure A.4: Patents in the KPSS vs DS Mappings to Compustat



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.

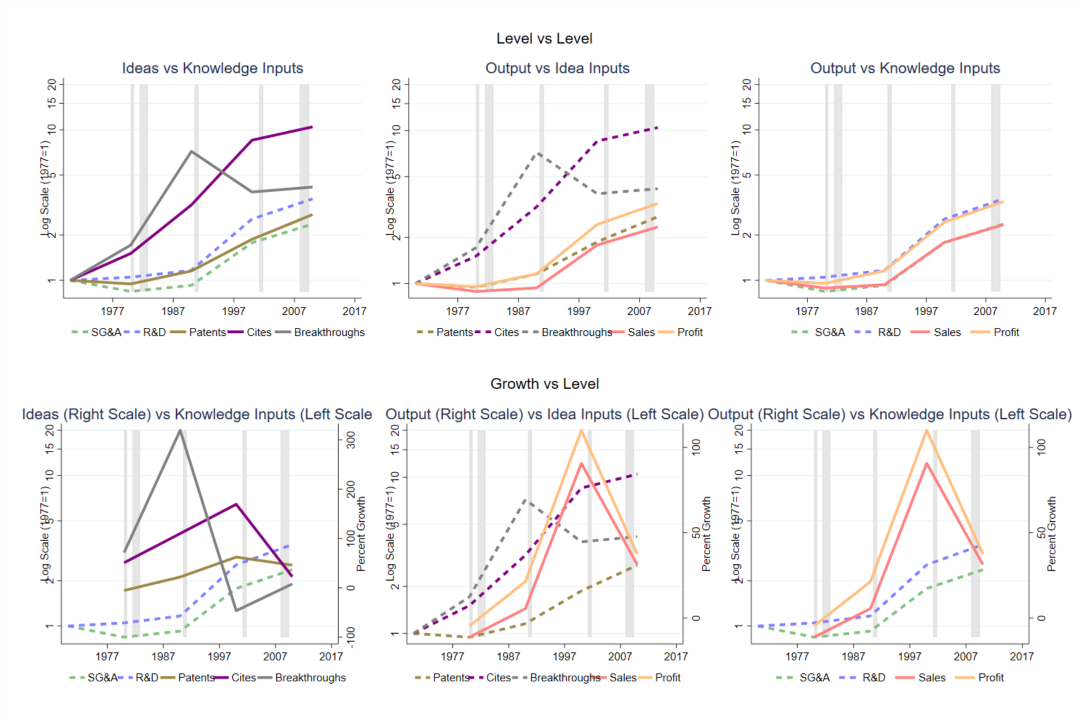
C Aggregate Compustat Sales and Operating Profit

Figure A.5: Aggregate Growth Rates (Compustat Panel)



Source: Compustat and authors' calculations. Figure displays the t- to 4-year aggregate growth rates for sales, operating profit and TFP across firms in Compustat.

Figure A.6: Knowledge Inputs, Ideas and Output in the Compustat Panel



Source: PV, KPSS, KPST, Compustat, DS, and authors' calculations. Figure displays aggregate knowledge inputs, patenting, sales and operating profit across firms in the Compustat panel. Top row compares levels to levels. Bottom row compares growth rates to levels.

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.[13]

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

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

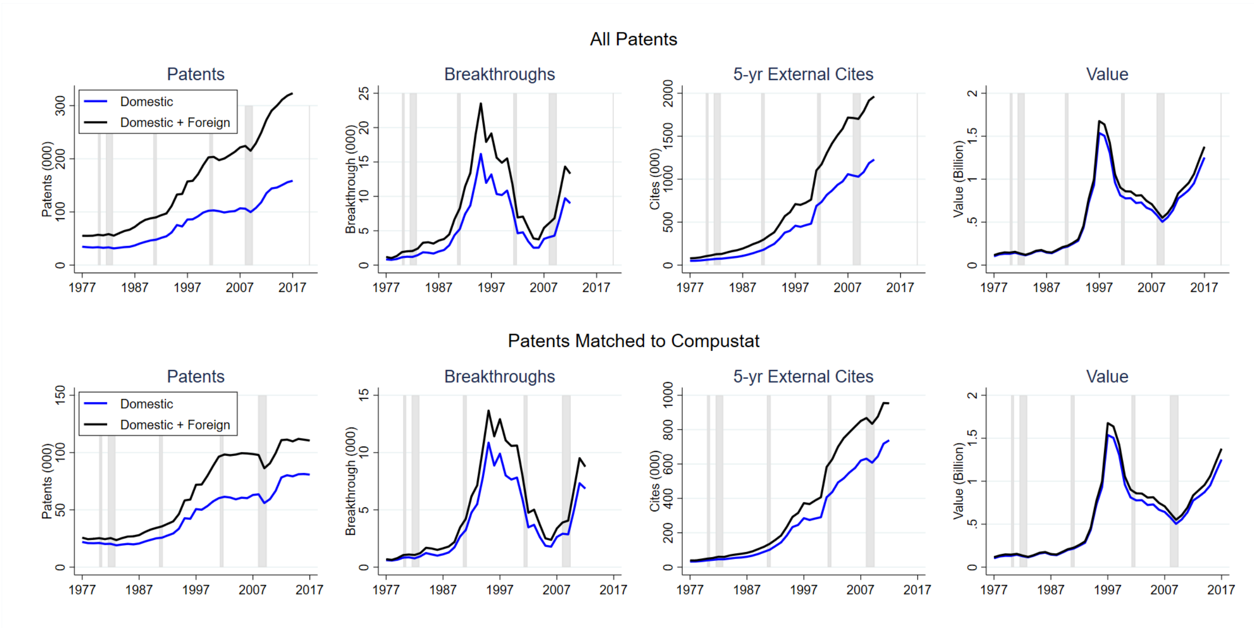
collapse EINs to the firm level and develop code to the match the data to our LBD panel. In practice, the enterprise identifiers in the RADs are not always EINs. We cycle over various Census identifiers and also match to a window of years to improve matching. This procedure is especially important in 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 Additional Patent-Level Information

The panels in the top row of Figure A.7 provide a breakdown of the share of patents, breakthrough patents, 5-year external citations and patent value attributable to "domestic" versus "foreign" patents in the raw PV dataset, where a domestic patent is defined as one in which at least one assignee has is located in the United States. The panels in the bottom row of Figure A.7 offer an analogous breakdown for the set of patents that match to Compustat.

Figure A.7: Domestic vs Foreign Patents by Application Year



Source: PV and authors' calculations. Figure provides a breakdown of the share of patents, breakthrough patents, 5-year external citations and patent value attributable to "domestic" versus "foreign" patents. A domestic patent is defined as one in which at least one assignee has is located in the United States. Top row is for the raw PV dataset. Bottom row is for patents matched to Compustat.

Table A.8 reports the number of patents, in thousands, of the top 25 patenting firms in the PV dataset from 1977 to 2017, as well as the percent of their patents that are "domestic", i.e., for which at least one assignee is located in the United States. Table A.9 reports analogous information for the top 25 patenting firms in Compustat over the same period. As indicated in both tables, the patents of U.S. based multinationals like IBM are almost exclusively domestic, while those of foreign multinationals like Canon or Sony are almost all foreign.

Table A.8: Top 25 Patenters in Raw PV Data, by Assignee 1977 to 2017

Assignee	Patents	Share Domestic
INTERNATIONAL BUSINESS MACHINES CORPORATION	137.1	99.97
SAMSUNG ELECTRONICS CO., LTD.	120.5	0.38
CANON KABUSHIKI KAISHA	73.8	0.38
SONY CORPORATION	54.7	6.66
FUJITSU LIMITED	52.9	0.49
KABUSHIKI KAISHA TOSHIBA	49.2	0.58
GENERAL ELECTRIC COMPANY	45.4	99.76
MITSUBISHI HEAVY INDUSTRIES, LTD.	42.7	0.36
HITACHI, LTD.	42.6	0.56
INTEL CORPORATION	41.6	99.80
SUMITOMO ELECTRIC INDUSTRIES, LTD.	36.5	0.47
LG ELECTRONICS INC.	34.0	0.07
NEC CORPORATION	31.5	0.38
MICROSOFT CORPORATION	30.1	99.95
SIEMENS AKTIENGESELLSCHAFT	28.1	2.14
MICRON TECHNOLOGY, INC.	27.0	99.93
SEIKO EPSON CORPORATION	26.8	0.28
TOYOTA JIDOSHA KABUSHIKI KAISHA	26.3	0.34
HONDA MOTOR CO., LTD.	25.2	0.94
RICOH COMPANY, LTD.	24.9	2.59
HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P.	24.5	99.93
QUALCOMM INCORPORATED	24.3	99.79
ROBERT BOSCH GMBH	24.2	2.17
TEXAS INSTRUMENTS INCORPORATED	23.3	99.97
SHARP KABUSHIKI KAISHA	23.1	0.93

Source: PV and authors' calculations. Table reports the number of patents, in thousands, of the top 25 patenting firms in the PV dataset from 1977 to 2017, as well as the percent of their patents that are "domestic", i.e., for which at least one assignee is located in the United States. Patents can have multiple assignees. For the purposes of this table, each assignee is credited with a patent.

Table A.9: Top 25 Patenters in Compustat Sample, by *permno* 1977 to 2017

permno	Name	Patents	Share Domestic	Break-throughs	Share Domestic	Citations	Share Domestic	Real Value	Share Domestic
12490	IBM	116.3	99.8	10.8	100.0	258.4	99.8	355.2	99.8
21152	CANON INC	75.7	0.8	4.2	2.0	131.1	1.3	21.6	1.6
51131	SONY	61.8	7.8	3.8	20.6	157.4	11.2	15.4	9.1
64231	HITACHI LTD	58.8	1.0	4.3	1.8	188.7	1.9	4.2	1.2
53727	PANASONIC	55.9	2.1	2.7	7.4	157.4	3.1	4.9	1.2
12060	GE AEROSPACE	52.0	94.8	1.5	97.7	102.4	96.9	838.4	96.9
59328	INTEL CORP	42.0	96.5	3.2	99.3	110.6	98.0	570.0	98.5
10107	MICROSOFT	41.0	99.1	6.0	99.8	183.5	99.1	624.3	99.7
27828	HP INC	35.1	98.6	3.7	99.8	108.8	99.8	299.5	99.4
59555	HONDA	32.5	1.4	0.8	1.8	73.3	1.9	9.8	0.8
76655	TOYOTA	28.2	8.0	0.7	24.2	54.1	7.1	10.2	9.1
55782	NEC CORP	27.3	2.5	3.4	4.5	103.7	4.5	0.6	1.7
53613	MICRON	26.2	95.2	1.8	99.9	67.8	96.3	88.9	97.9
37867	FUJIFILM	25.3	0.6	1.2	0.1	54.9	0.5	5.7	0.0
77178	QUALCOMM INC	25.2	99.4	1.9	99.6	62.0	99.5	195.7	99.4
15579	TEXAS INSTR	22.9	97.7	2.3	98.7	80.2	97.5	226.6	96.6
17830	RTX CORP	21.1	91.8	0.3	97.5	30.8	94.2	144.9	99.8
22779	MOTOROLA	20.9	99.6	4.7	99.8	117.4	99.3	177.6	99.7
10145	HONEYWELL	20.2	96.2	0.7	98.8	58.7	98.0	133.3	98.7
88935	SIEMENS AG	20.2	27.6	0.4	39.0	36.5	37.9	5.5	30.6
88487	PHILIPS	19.8	4.4	0.5	5.3	54.4	13.0	7.8	3.3
27983	XEROX	19.2	99.6	1.8	99.9	46.1	99.7	59.5	99.5
33452	ERICSSON	18.8	12.0	4.1	27.7	69.6	31.8	40.4	36.5
14593	APPLE INC	17.6	99.8	0.8	100.0	71.6	99.8	569.6	99.9
11754	KODAK	16.5	99.3	0.8	98.6	42.6	99.3	55.8	99.7

Source: Compustat, PV, KPSS, DS, and authors' calculations. Table reports the number of patents, breakthroughs, 5-year external citations and patent value from 1977 to 2017 among the top 25 patenting firms, as well as the percent that are that are from "domestic" patents, i.e., those for which at least one assignee is located in the United States. Patents, breakthroughs and citations are expressed in thousands. Value is expressed in billions of 1982 dollars.

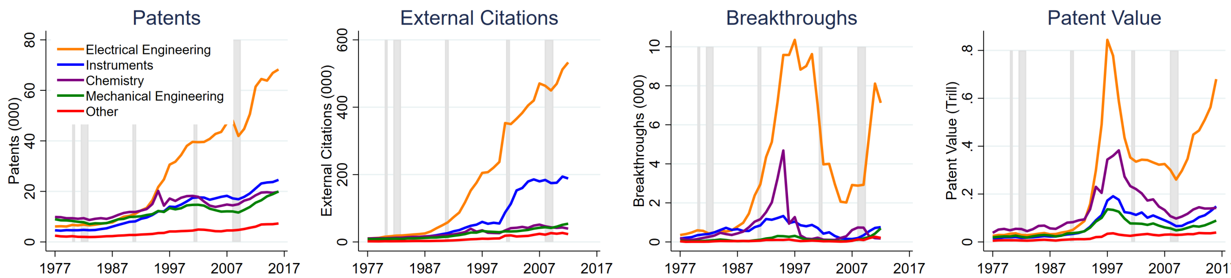
F Patent Types Over time

Figure A.8 presents a breakdown of US patenting activity by World Intellectual Property Organization (WIPO) categories, as recorded in PV.[14] 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 1 is largely driven by innovation in Electrical Engineering, which includes patents related to computing, telecommunications, information technology, and semi-conductors. This growth occurs in two bursts starting in the late 1980s and after the Great Recession.[15]

[14] While Figure A.8 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.

[15] 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.

Figure A.8: Patenting Activity by WIPO Code



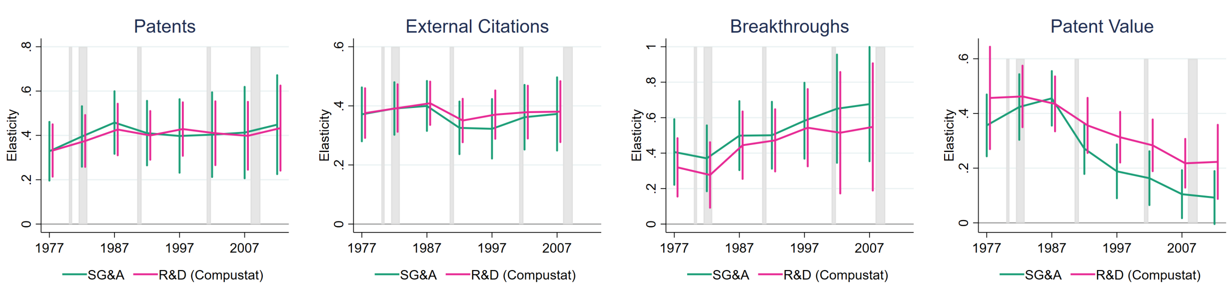
Source: PV, KPSS, KPST and authors' calculations. Notes: Figure provides a breakdown of patents, external citations, breakthrough patents and patent value by WIPO category and year. Patent value is in trillions of 1982 dollars.

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.

G Additional Patent Efficiency Estimates

Figure A.9 compares the baseline patent elasticities from Figure 8 in the main text to the results of an alternate estimation where counts of patents, breakthroughs and citations are with respect only to the patents for which value can be observed.

Figure A.9: Elasticities Restricted to Patents with Value



Source: Compustat, PV, KPSS, DS, KPST, and authors' calculations. Figure reports comparison of the 95 percent confidence intervals for patent efficiency coefficients (β_s) for the baseline specification reported in the main text to an alternate specification where counts of patents, breakthroughs and citations are with respect only to the patents for which value can be observed. Each regression is restricted to the set of firms with positive SG&A or R&D in each year, whether or not they patent. Granted patents are assigned to firms in their application year.

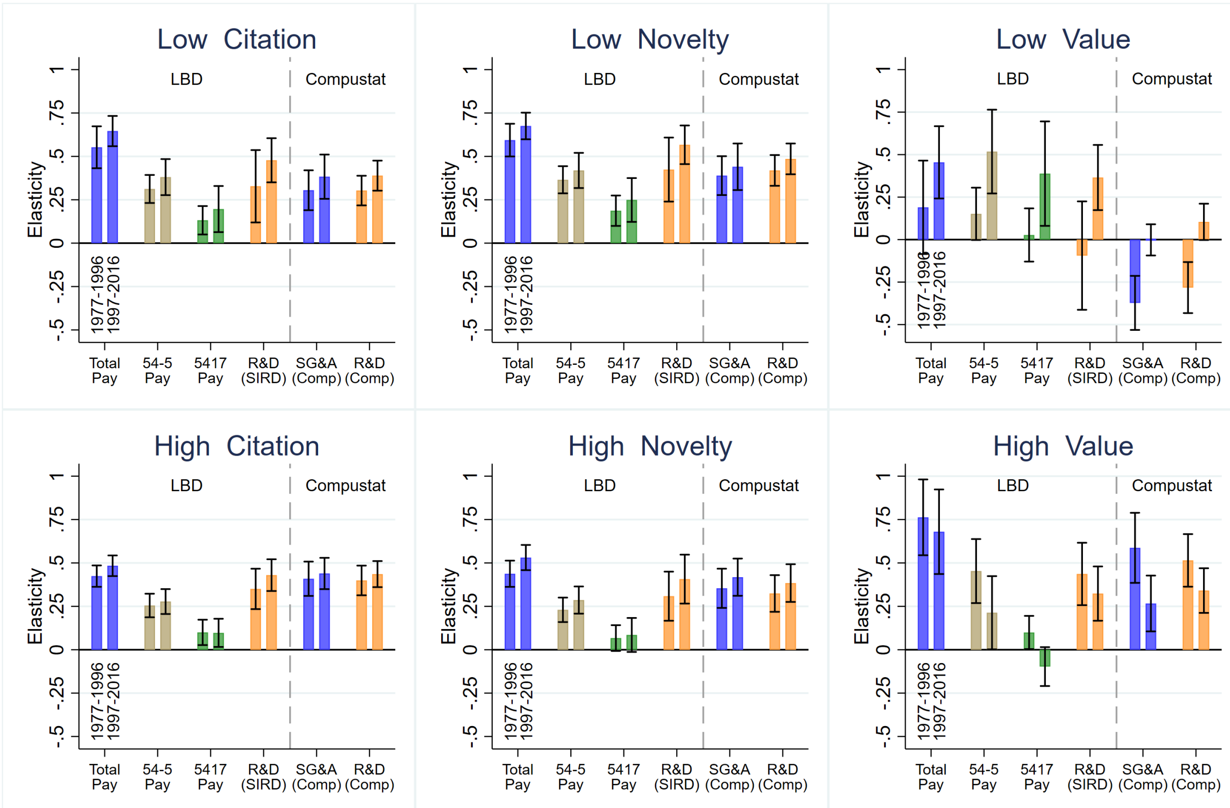
H Pre vs Post Patent Efficiency Estimates

We investigate the relationship between knowledge inputs and patent outcomes further by estimating separate elasticities for *counts* of low- versus high-value patents for each quality measure in each year of the sample. We define low-quality patents as those in the first quartile in that year, while high-quality are those in the fourth quartile.[16] This approach ensures that our left-hand side variables are comparable counts across all measures of patenting activity, and exploits variation in the firms and their timing of input-use and 'high' versus low-value outcomes. To reduce disclosure burden, we estimate just two elasticities for each count and knowledge stock, the first for the years leading up to and including 1996 and the second for years 1997 and beyond. Figure A.10 plots these estimates. The first 8 bars in each panel are for the 4 knowledge stocks computed for the LBD panel, while the remaining 4 bars, separated by the vertical dashed line, represent the estimated elasticities for SG&A and R&D stocks in Compustat.

Figure A.10 shows that average patent elasticities rise over time for both low- and high-citation patents and for both low- and high-novelty patents. Elasticities also rise over time for low-value patents. By contrast, estimated elasticities fall over time for high-value patents. We find a similar message using the Compustat panel. Firms seem to be improving in their ability to translate R&D inputs into cited and novel ideas, even as their ability to generate ideas that the stock market predicts will lead to future growth declines.

[16]This separation is straightforward for patent value and citations, though we note that more than one quarter of patents in each year typically have zero external citations; we choose among those zeros randomly. As breakthroughs are defined across-time, we define patents to be low- or high-novelty if their underlying 5-year backward versus forward textual similarity ratios is in the first versus fourth quartile in each year.

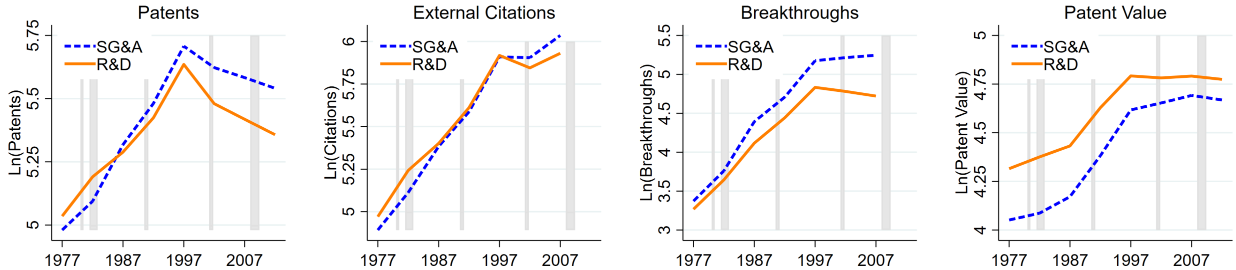
Figure A.10: Low- vs High-Quality Patent Elasticities Before and After 1996



Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, Compustat, and authors' calculations. Notes: Figure reports estimates of η_j by knowledge stock for the LBD and Compustat panels on a version of Equation 6 where the left-hand side variables are the *counts* of low- or high-quality patents as defined in the text and the right-hand-side dummies pick out just two periods, the years leading up to and including 1996 and the years after 1997. Whiskers correspond to 95 percent confidence intervals. Standard errors are clustered at the firm-level.

I Estimated Patent Elasticity Firm Fixed Effects (γ_j)

Figure A.11: Aggregate Estimated Firm Fixed Effects (γ_f) by Semi-Decade

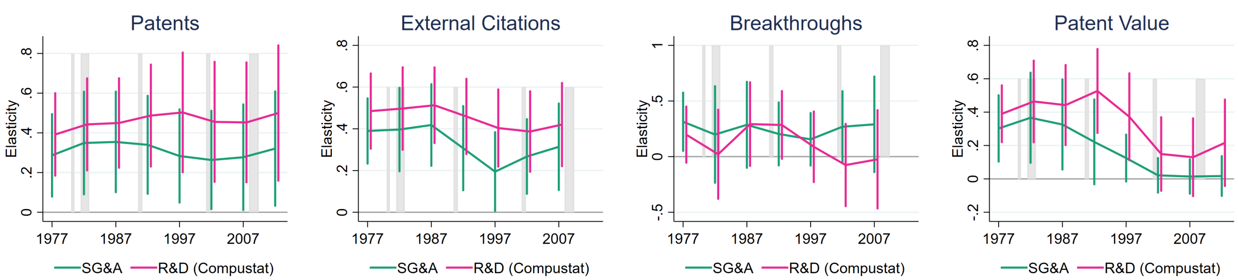


Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, Compustat and authors' calculations. Notes: Figure reports the log sum of firms' exponentiated fixed effects, i.e., $ln\left(\sum_f e^{\gamma_f}\right)$, from Equation 6 by patenting activity and knowledge stock for the LBD and Compustat panels.

J Estimates for Balanced Panel of Compustat Firms

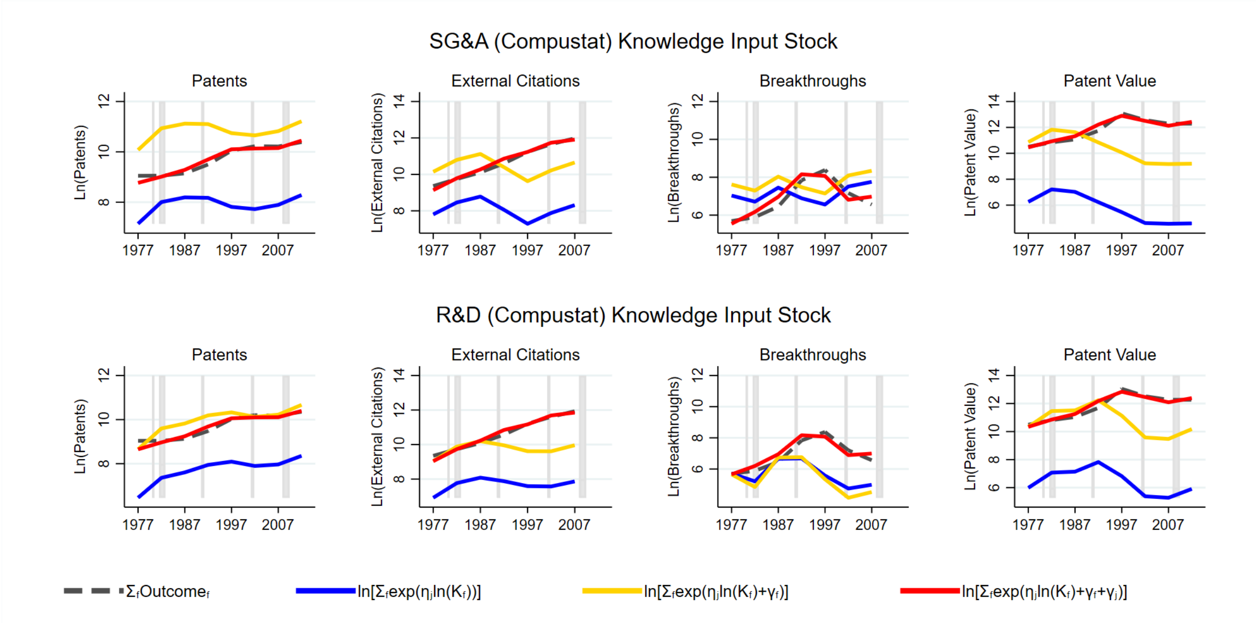
This section reports versions of Figures 8 and A.13 using estimates or Equation 6 from the balanced panel of Compustat firms.

Figure A.12: Estimated Patent Elasticities by Semi-Decade (η_j) for Balanced Compustat Panel



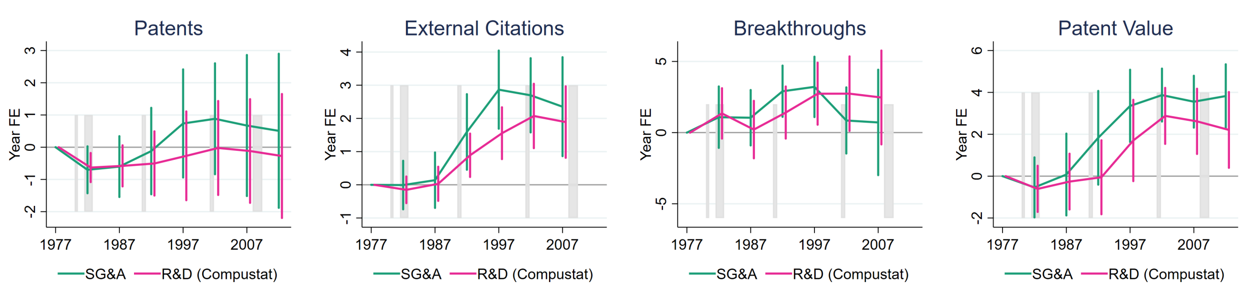
Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, Compustat and authors' calculations. Notes: Figure reports estimates of η_j's from Equation 6 by patenting activity and knowledge stock for the LBD and Compustat panels. Whiskers denote 95 percent confidence intervals. Standard errors are clustered at the firm-level.

Figure A.13: Predicted Growth in Patenting from Elasticity Estimates for a Balanced Compustat Panel



Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, Compustat and authors' calculations. Notes: Figure reports growth in patents, external citations, breakthrough patents and patent value over the 1977 to 2007 (external citations, breakthroughs) or 2012 (patents, patent value) sample period.

Figure A.14: Estimated Semi-Decade Fixed Effects (γ_j) for a Balanced Compustat Panel



Source: PV, LBD, KPSS, KPST, RADS, BDSPF-Long, Compustat and authors' calculations. Notes: Figure reports estimates of γ_j from Equation 6 by patenting activity and knowledge stock for the LBD and Compustat panels. Whiskers denote 95 percent confidence intervals. Standard errors are clustered at the firm-level.

K Replication of KPSS Tables IV and V

KPSS Tables IV and V present their main results linking firm growth to own and competitor patent values and own and competitor citations. Their firm-level OLS estimating equation (equation 12 in the paper) is:

$$logX_{f,t+\tau} - logX_{f,t} = a_\tau \theta_{f,t} + b_\tau \theta_{I\setminus f,t} + cZ_{ft} + u_{ft+\tau}. \tag{A.1}$$

The sample period is 1950 to 2010.KPSS consider $\tau \in (0,4)$. The left-hand side is the log growth of outcome X between years t and $t+\tau$. The first and second terms on the RHS are firm f's patent ($flow$) value in year t (θ_{ft}) and the total ($flow$) value of patents by other firms in the same 3-digit SIC industry in year t ($\theta_{I\setminus f,t}$). These variables are both scaled by total assets, i.e., own patent value per own assets and competitor patent value per competitors' assets, respectively. The third term on the RHS is a set of controls encompassing the lagged dependent variable and lagged capital, labor and a measure of firms' stock volatility.

KPSS report their results in terms of "beta coefficients", i.e., multiplying the estimate coefficients of interest by their standard deviation. In Table A.10 below, I report my replication of their Table IV. Comparison of this table with the one in the paper reveals that the coefficients match well, the t-stats a bit less so in terms of magnitude but not necessarily significance. The latter difference could be do to my use of $reghdfe$ versus their use of reg.As indicated in Table A.10, across all outcomes, own patent value is associated with positive growth, while competitor patent value is linked to lower growth.

Table A.10: Replication of KPSS Table IV: Outomes and Patent Value

Own Value					Competitor Value				
0	1	2	3	4	0	1	2	3	4
Panel A Profits									
0.018	0.029	0.036	0.043	0.047	-0.016	-0.031	-0.035	-0.042	-0.050
3.057	3.865	3.387	3.409	3.251	-3.566	-6.075	-6.788	-6.489	-6.832
Panel B Output									
0.008	0.014	0.020	0.025	0.033	-0.017	-0.034	-0.043	-0.050	-0.059
2.428	2.588	2.604	2.461	3.012	-5.750	-8.452	-7.952	-7.664	-7.384
Panel C Capital									
0.010	0.019	0.027	0.033	0.038	0.002	-0.006	-0.014	-0.023	-0.033
7.332	6.031	4.862	3.806	3.492	0.437	-0.957	-1.933	-2.647	-3.488
Panel D Labor									
0.006	0.011	0.016	0.020	0.022	-0.005	-0.015	-0.019	-0.021	-0.023
4.278	3.484	3.288	3.041	2.695	-1.363	-3.225	-4.011	-3.398	-3.276
Panel E TFPR									
0.013	0.017	0.019	0.023	0.025	-0.003	-0.007	-0.009	-0.014	-0.018
2.020	2.033	2.455	3.167	3.965	-1.428	-2.478	-2.938	-3.947	-4.764

Source: Compustat, PV, KPSS. Table reports replication of KPSS Table IV. Each pair of rows in the table contain estimates from a separate regression. The first row in each pair reports the "beta coefficients" while the second row reports the t-stats. Left panet reports results for own patent value (a_τ) while right panel reports results for competitors' patent value (b_τ).

To demonstrate the value of their measure over existing estimates of patent value, KPSS compare the results for growth as a function of own and competitor patent value to results from estimation of a similar specification that uses own and competitor citations instead of patent value. They report these results in their Table V. Table A.11 reports my replication of that table, using their citation measure. Here, too, my coefficient estimates line up well with those reported in the KPSS paper. One thing to note regarding these results is that while *own* citations do a relatively good job with TFP, *competitor* citations do not show much of a relationship with TFP, $Profit$ or $Output$.

Table A.11: Replication of KPSS Table V: Outcomes and Citations

Own Value					Competitor Value				
0	1	2	3	4	0	1	2	3	4
Panel A Profits									
0.006	0.011	0.017	0.023	0.029	-0.002	0.000	0.000	0.005	0.004
3.709	4.354	4.218	4.558	5.002	-0.773	0.017	0.076	0.789	0.644
Panel B Output									
0.001	0.002	0.008	0.013	0.017	0.000	0.002	0.004	0.005	0.007
0.444	0.786	2.392	2.501	2.881	0.022	0.555	0.910	0.977	1.063
Panel C Capital									
-0.005	-0.004	-0.001	0.005	0.008	0.003	0.005	0.006	0.008	0.010
-3.816	-1.758	-0.230	1.021	1.501	1.771	1.706	1.486	1.582	1.732
Panel D Labor									
-0.004	-0.003	0.002	0.005	0.007	0.006	0.010	0.013	0.016	0.019
-2.705	-1.076	0.486	1.293	1.463	2.766	3.312	3.633	3.295	3.282
Panel E TFPR									
0.005	0.008	0.008	0.010	0.010	-0.001	-0.001	0.002	0.003	0.004
3.661	4.368	3.921	4.057	3.834	-0.951	-0.347	0.772	1.068	1.186

Source: Compustat, PV, KPSS. Table reports replication of KPSS Table V. Each pair of rows in the table contain estimates from a separate regression. The first row in each pair reports the "beta coefficients" while the second row reports the t-stats. Left panet reports results for own patent value (a_τ) while right panel reports results for competitors' patent value (b_τ).