Technology Differentiation and Firm Performance

Sam Arts
Bruno Cassiman
Jianan Hou

Working Paper 22-040

Technology Differentiation and Firm Performance

Sam Arts
KU Leuven

Bruno Cassiman
Harvard Business School

Jianan Hou
KU Leuven

Working Paper 22-040

Funding for this research was provided in part by Harvard Business School. Sam Arts acknowledges support from KU Leuven (grant 3H200208) and the National Bank of Belgium (grant 3H160338). Bruno Cassiman acknowledges support from FWO grant G071417N from the Flemish government.

Technology Differentiation and Firm Performance[1]

Sam Arts
Department of Management, Strategy and Innovation
Faculty of Economics and Business
KU Leuven
sam.arts@kuleuven.be

Bruno Cassiman
Harvard Business School[2]
bcassiman@hbs.edu

Jianan Hou
Department of Management, Strategy and Innovation
Faculty of Economics and Business
KU Leuven
jianan.hou@kuleuven.be

December 2021

ABSTRACT

Prior work has extensively studied how investing in R&D and building a technology portfolio relate to superior firm performance. However, the value of a firm's technology portfolio should also be driven by the degree to which it is more unique and technologically differentiated from other firms. To study this research question, we develop a new method to characterize firm technology based on the semantic content of patent portfolios that allows us to map a firm's competitive position in the technology space relative to all other firms and to measure the differentiation of a firm's technology portfolio. Using a large panel of U.S. public firms from 1980 to 2015, we find that technology differentiation has a strong positive and long-lasting relation with firm performance. Moreover, differentiated firm technology is particularly valuable in industries with higher R&D intensity and with stronger product market competition. We provide open access to all code and data to measure the technology similarity and the technology differentiation of U.S. public firms.

[1] Acknowledgement: Comments from Juan Alcacer, Paul Anckaert, Andrea Fosfuri, Tim Simcoe, Giovani Valentini and seminar participants at Boston University, Cornell University, Harvard Business School, NBER, Tsing Hua University, Utrecht University and the 2021 EPIP and Druid conferences improved the paper substantially. Sam Arts acknowledges support from KU Leuven (grant 3H200208) and the National Bank of Belgium (grant 3H160338). Bruno Cassiman acknowledges support from FWO grant G071417N from the Flemish government.

[2] On leave from Department of Management, Strategy and Innovation, Faculty of Economics and Business, KU Leuven and IESE Business School.

1. Introduction

According to the resource-based view, a firm's competitive advantage and superior performance relies on resources that are unique and difficult to substitute or imitate, such as unique and proprietary firm technology (Wernerfelt 1984, Peteraf 1993, Mowery et al. 1998). Prior empirical studies show how R&D investments and the size of a firm's technology portfolio as measured by patents positively relate to firm performance (Hall et al. 2005, Simeth and Cincera 2016, Bellstam et al. 2021).[3] But, the value of a firm's technology portfolio should also be driven by its uniqueness and the degree to which it technologically differentiates from all other firms. A limited number of formal theory models show that firms tend to differentiate their technology from other firms to minimize technology spillovers and competition in the product market (Kamien and Zang 2000, Gil Moltó et al. 2005, Aghion et al. 2005, Lin and Zhou, 2013).

In this paper, we empirically study how technology differentiation relates to firm performance. To do so, we develop a new method to characterize firm technology based on the semantic content of patent portfolios that allows us to map each firm's competitive position in the technology space relative to all other firms and to measure the uniqueness or differentiation of a firm's technology portfolio. To characterize firm technology portfolios, prior work has relied on patent classification or citations (e.g. Jaffe 1989, Rosenkopf and Almeida 2003, Bloom et al. 2013, Stuart and Podolny 1996). However, this approach has well-known limitations. Patent citations only capture prior art and not the content of a firm's own technology. In contrast to citations, patent classification does reflect subject matter but is generally too broad to capture the detailed content of a firm's technology and measure technology differentiation (Thompson and Fox-Kean 2005). Instead of relying on structured patent information, we exploit the fact that firms must provide a fully written disclosure of their technological inventions in exchange for legal protection. Compared to the traditional approach based on patent classification or citations, patent text provides a more detailed insight into the technology portfolio of firms (Arts et al. 2018, Righi and Simcoe 2019), and particularly outperforms in the identification of pioneering technologies, such as the polymerase chain reaction, the FinFET transistor, or the lithium-ion battery (Arts et al. 2021a). Identifying pioneering technologies is important to assess whether a firm is pushing the technology frontier and to measure the firm's degree of technology differentiation relative to other firms.

Our data collection draws from the DISCERN patent database,[4] which dynamically matches public U.S. firms to U.S. patents for the period 1980-2015 (Arora et al. 2021), and from the processed and cleaned text of U.S. patents from Arts et al. (2021a). Representing each firm-year level patent portfolio as

[3] Not all technologies are patented (Hall et al. 2014). Nevertheless, in line with prior work we rely on firm patent portfolios to characterize a firm's technology portfolio.
[4] DISCERN database is available from: https://zenodo.org/record/3709084.

a vector in which each dimension corresponds to one stemmed technical keyword, we calculate cosine similarities with term frequency–inverse document frequency weights (TF-IDF) to measure the pairwise technology similarity between all U.S. public firms for all years and use these to construct our measure of technology differentiation for each firm and year. Our sample includes 4,832 firms, 57,772 firm-year level patent portfolios, and 98,279,118 pairwise technology similarities. As expected, younger, smaller and more R&D intensive firms display higher levels of technology differentiation relative to older, larger and less R&D intensive firms.

Next, we study the relation between our new firm-year level measure of technology differentiation and firm performance. Using firm fixed effects models on the entire panel of U.S. public firms for the years 1989-2015, we find that technology differentiation has a strong positive and long-lasting relation with firm profitability (return on assets) and market value (Tobin's Q). Controlling for a firm's R&D intensity and the number of citation-weighted patents in the technology portfolio, one standard deviation increase in technology differentiation corresponds with an increase of 13.4% in Tobin's Q and 3.3% in ROA. In addition, we find that unique and differentiated firm technology is particularly valuable in R&D intensive industries and in industries with strong product market competition. Finally, when technology differentiation is calculated based on the traditional characterization of firm technology portfolios using either patent classification or citations there is no significant relationship with firm performance. As we argue that unique and differentiated technology relates to a competitive advantage and superior firm performance, our findings suggest that a text-based measure more accurately characterizes a firm's technology portfolio to map a firm's competitive position and differentiation in technology space relative to other firms.

Our paper makes several contributions to the strategy and innovation literature. First, we introduce a new method and data to characterize firm technology portfolios that allows us to map each firm's competitive position in the technology space relative to all other firms and to measure the uniqueness or differentiation of a firm's technology over time. We demonstrate that our method provides a different characterization of firm technology portfolios to the traditional approach based on patent classification or citations, and that our method more strongly correlates with firm profitability and market value. Although our analysis is restricted to U.S. public firms, we illustrate that our method also works for firms with only a few patents in their portfolio. As such, our method could presumably also be used to characterize the technology portfolio and measure the competitive position and differentiation in the technology space of smaller (non-public) firms and startups. In addition, we show that the method works for firms specialized in a single product market industry as well as for diversified firms operating in multiple industries. Second, whereas the economics and strategy literature has predominantly focused on a firm's competitive position and differentiation in the product market (e.g. Hoberg and Phillips 2016, Guzman and Li 2019), we introduce and test the importance of a firm's competitive position and differentiation in the technology space. Using a large sample of U.S. public firms linked to patents, we show that technology differentiation has a strong and robust relationship with firm performance, and that unique and differentiated firm technology is particularly valuable in industries with higher R&D intensity and with stronger product market competition. As a result, our contribution paves the way to explore different mechanisms that relate technology differentiation to competitive advantage and firm performance, and to study different drivers of technology differentiation. Finally, we provide open access to our code and data that might be useful to visually represent a company's competitive position in the technology space (Stuart and Podolny 1996), to cluster technologically similar firms (Jaffe 1989), to identify technology pioneering firms with unique and differentiated technology (Ahuja and Lampert 2001), to more accurately measure technology spillovers between firms (Bloom et al. 2013), to assess the potential for technology synergies in case of M&As or alliances (Rosenkopf and Almeida 2003), or to study the evolution and diversification of firms' technology strategies (Silverman 1999).

2. Methodology

2.1 Data and sample

To obtain firm patent portfolios, we rely on the DISCERN database that dynamically matches public U.S. firms to U.S. patents for the period 1980-2015 (Arora et al. 2021).[5] We start from a sample of 1,345,945 U.S. patents granted between 1980 and 2015 that are ultimately assigned to any public U.S. firm. In order to construct the patent portfolio of firm *i* in year *t*, we collect all patents owned by firm *i* with a filing year between year *t-5* and year *t-1* (Ahuja and Katila 2001, Rothaermel and Deeds 2004, Hirshleifer et al. 2018). For each firm, we obtain additional information from Compustat and data on product market similarities between firms from Hoberg and Phillips (2016). The final sample includes 4,832 firms and 57,772 firm-year level patent portfolios for the period 1980 to 2015. Next, we use the processed, cleaned, and stemmed technical keywords extracted from the titles, abstracts, and claims of each U.S. patent from Arts et al. (2021).[6]

---

[5] DISCERN dynamically matches each patent assignee to its ultimate owner in different time periods by combining multiple sources such as Orbis, SDC Platinum, CRSP monthly stock, and the NBER patent database in order to account for firm name changes, subsidiaries, and ownership changes because of M&As.

[6] For each patent, they concatenate title, abstract, and claims, lowercase text and tokenize it to words using the following regular expression: [a-z0-9][a-z0-9-]*[a-z0-9]+|[a-z0-9]. They consider a word as a sequence of letters and numbers that could be separated by hyphens ("-"). Next, they remove words composed only by numbers, one-character words, stop words from the Natural Language Toolkit (NLTK) in the Python library, and words appearing in only one patent. In addition to natural stop words, they remove a manually compiled list of 32,255 very common non-technical keywords. Finally, they apply stemming to each word using the SnowBall method.

2.2 Characterizing firm technology portfolios and measuring technology similarity

To map each firm's competitive position in the technology space and measure technology differentiation, we represent the patent portfolio of firm *i* in year *t* as a vector of 1,030,335 dimensions where each dimension corresponds to one stemmed technical keyword from the entire patent vocabulary, and each entry captures the share of patents from firm *i*'s patent portfolio in year *t* that contain the particular keyword. To illustrate this approach, Figure 1 displays word clouds based on the 100 most frequent stemmed technical keywords in the patent portfolio of four different companies. For Tesla, some of the most common stemmed keywords include *batteri*, *pack*, *vehicl*, and *charg*; for Monsanto *transgen*, *pollen*, *herbicid*, and *glyphos*; for 3M *adhes*, *polymer*, *substrat* and *alkyl*; and for Universal Display Corporation (manufacturer of organic light-emitting diodes) *emit*, *cathod*, *anod*, and *phosphoresc*. The average firm-year level patent portfolio in our sample includes 1,658 stemmed keywords providing a detailed insight in the firm's technology portfolio.[7]

'Insert Figure 1'

Using these firm-year level vectors representing the technology portfolio of a firm at a point in time (n=57,772), we compute for each year *tech similarity* for every pair of firms by means of cosine similarities (see also Jaffe 1986, 1989). We use term frequency–inverse document frequency (TF-IDF) weights. The frequency of a keyword in a particular firm-year patent portfolio, i.e. the share of patents containing the keyword, is then offset by the share of all firm patent portfolios from the entire population in a given year which contain the particular keyword.[8] This helps to adjust for the fact that some keywords are more representative of a firm's patent portfolio (e.g. *batteri* for Tesla or *herbicid* for Monsanto) and for the fact that some keywords appear frequently across all patent portfolios and are therefore less discriminating across firms (e.g. *electr* or *drug*). Moreover, pioneering technologies as measured by new (or recent) keywords introduced for the first time in history and which occur less frequently across all patent portfolios receive a higher weight (Arts et al. 2021a). We calculate for each year the *tech similarity* for every pair of firms. The dataset covers the years 1980-2015 and includes 4,832 firms, 57,772 firm-year observations,[9] and 98,279,118 pairwise *tech similarities* for all firm-year observations. Our new text-based

[7] For instance, we retrieved 22,743 stemmed keywords from 6,895 Monsanto patents, 41,078 from 15,237 3M patents, 2,020 from 227 Tesla patents, and 2,989 from 299 Universal Display patents.

[8] The TF-IDF adjustment is conducted as follows. First, we construct a vector for each firm-year where the value of each dimension captures the term frequency (TF), namely the share of patents in the firm portfolio using the given word. Second, for each word we calculate its inverse document frequency (IDF), namely the total number of firms in the sample divided by the number of firms using this given word. Due to the high skewness of IDF, we take the logarithm with base 10. Finally, the adjusted value equals the product of TF and IDF. The TF-IDF weights are calculated for each firm-year observation.

[9] A firm-year observation is only included in case the firm has at least one patent in its portfolio (i.e. in years t-5 to t-1).

tech similarity between firms is only moderately correlated with the traditional similarity measures based on patent classification or citations (e.g. Jaffe 1989).[10] As an illustration, Table A.2 in the Appendix ranks, for a selected number of focal firms, the top 10 most similar firms according to different measures. There is only a moderate overlap between the firms identified by *tech similarity* and those identified by the traditional metrics. Thus, the semantic content of patent portfolios provides a different characterization of firm technology and a firm's competitive position in the technology space compared to prior established approaches.[11] Interestingly, the correlation between *tech similarity* and the product similarity between firms, based on the business descriptions from annual 10K reports, is only 0.25 (Hoberg and Phillips 2016). [12] This suggests that companies competing with similar products in the same industry often rely on different types of technology. The companies with the most similar technology portfolios in history include IBM and Digital Equipment in 1994 (*tech similarity*=0.904), Baker Hughes and Schlumberger (both providing oil field services) in 2004 (*tech similarity*=0.906), AT&T and Sprint in 2006 (*tech similarity*=0.906), Alphabet (Google) and Altaba (Yahoo!) in 2009 (*tech similarity*=0.931), and Texas Instruments and Freescale Semiconductor in 2012 (*tech similarity*=0.923).

Our open access code and data can be used to map and visualize a company's competitive position in the technology space relative to all – or to a selected number of – other firms, and to cluster firms based on technology similarity, in line with the text-based product industry clustering of Hoberg and Phillips (2016). Scholars can use our input data together with their preferred clustering and visualization methods such as k-means or hierarchical clustering. As an illustration, Figure 2 shows a network graph restricted to all firms in the machinery industry in 2005. Each node represents one firm, the size of the node is proportional to the size of the firm's patent portfolio in 2005 (based on patents from 2000-2004), two nodes are connected by an edge in case *tech similarity* between the firms is above 0.6, and the thickness of the edge is proportional to *tech similarity* between the firms (thicker edge means higher *tech similarity*). The colors represent six clusters of technologically similar firms based on hierarchical clustering using the Ward method: semiconductor equipment firms (e.g. Applied Materials), engine technology firms (e.g. Caterpillar), fluid and air filter technology firms (e.g. Donaldson Company), mining, oil, and gas drilling

[10] In order to compare our new text-based *tech similarity* measure with traditional metrics, we calculated *tech similarity (class)*, *tech similarity (subclass)*, and *tech similarity (citation)* in the exact same way except for using all of a patent's classes, subclasses, and backward patent citations instead of keywords to characterize firm patent portfolios. Table A.1 in the Appendix provides a detailed overview of how each measure is calculated. The new text-based *tech similarity* measure has only a moderate correlation with *tech similarity (class)* (corr=0.38), *tech similarity (subclass)* (corr=0.35), and *tech similarity (citation)* (corr=0.22).

[11] Technological language might evolve over time. However, we are comparing firms at the same point in time and will control for time fixed effects in the analyses.

[12] We downloaded product similarity scores between 1988 and 2014 from the TNIC database (Hoberg and Phillips 2016). The similarity is only available for firm-years whose 10K filings are available and are covered by Compustat. Financial firms (SIC 6000-6999) and firm-years with nonpositive sales or with assets less than 1 million are excluded.

technology firms (e.g. Baker Hughes), photonics and laser technology firms (e.g. Veeco Instruments), and power tools and component technology firms (e.g. Black & Decker).[13,14]

'Insert Figure 2'

2.3 Technology differentiation

To measure the uniqueness and differentiation of firm *i*'s technology portfolio in year *t*, we calculate $tech\ differentiation_{it} = 1 - \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} tech\ similarity_{\ ijt}$, with *n* equal to all firms active in year *t* and $tech\ similarity_{ijt}$ equal to the technology similarity between firm *i* and firm *j* in year *t*.[15,16] Thus, firms with a more unique and less overlapping technology portfolio have a higher level of *tech differentiation*. Firms can increase their *tech differentiation* by moving away from crowded areas in technology space and focusing more on less crowded areas or emerging technologies, especially relative to the firm's closest competitors in technology space. Because of the TF-IDF weights, new or recent keywords capturing pioneering or emerging technologies receive a higher weight in the calculation of *tech differentiation* while older and more established keywords, which are very common in the portfolio of many companies, receive a lower weight. The correlation between *tech differentiation* of firm *i* in year *t* and the number of new keywords pioneered by firm *i* in the same period is 0.18, illustrating how firms pioneering new technologies increase their technology differentiation.[17] For instance, Merck invented – and pioneered the keyword – *gellan* (a biogum produced by bacteria which is widely used as thickener, emulsifier, and stabilizer) in 1983, increasing Merck's technology differentiation at that time. In 1996 Abbott Laboratories discovered – and was the first to introduce the keyword – *ritonavir*, a drug which dramatically reduced the death rate of HIV/AIDS. Texas Instruments pioneered the keyword *text-to-speech* (a speech computer or synthesizer producing human speech) for the first time in 1990. McAfee became more differentiated by being the first to introduce the keyword *malware* in a patent in 2000. As a final example, Overture Services filed a patent in 2003 for "a method and system for optimum placement

---

[13] As an illustration, Figure A.1 in Appendix shows a network graph for the same firms based on *tech similarity (class)* and illustrates once more the big differences in using patent text versus patent technology classification to map a firm's competitive position in technology space.

[14] As another example, Figure A.2 in Appendix shows the network of the top 100 firms with the largest patent portfolio in 2000 across all industries.

[15] As illustrated later, our main findings remain robust if we calculate *tech differentiation* of firm *i* in year *t* exclusively based on *tech similarity* of firm *i* in year *t* with all other firms from the same industry (3-digit SIC) in year *t* (instead of all firms from all industries), or if we calculate *tech differentiation* of firm *i* in year *t* as one minus the maximum (rather than the average) *tech similarity* between firm *i* and all other firms in year *t*.

[16] We also provide open access to the dataset covering the yearly technology differentiation measures for all 4,832 US public firms matched to patents between 1980 and 2015.

[17] The correlation is significant at the 1% level. *tech differentiation* also positively correlates with the share of patents from the firm's patent portfolio without any backward prior art citations (corr= 0.10, significant at 1% level), a measure which has been used in prior studies to identify pioneering technologies (Ahuja and Lampert 2001).

of advertisements on a webpage," the first patent pioneering the keyword *cost-per-click*, i.e. the internet advertising technology used by Google and others. The company was acquired by Yahoo! in the same year. The latter example illustrates how firms can increase their technology differentiation not only from internal R&D but also by acquiring external technology through M&As (Arora and Gambardella 1990; Arora et al. 2001). Therefore, the uniqueness and differentiation of technology portfolios might be a key driver of M&A transactions, particularly between product market rivals, that has been overlooked by prior empirical studies on the role of innovation for M&As (Arts et al. 2021b).

Tech differentiation can vary between industries, between firms within the same industry, and within a firm over time. As displayed in the variance decomposition in Table A.3 in Appendix, differences between industries account for 27% of the total variance in *tech differentiation*, differences between firms within the same industry account for 61% of the variance, and variation in *tech differentiation* within the same firm over time accounts for the remaining 12%. As expected, and as illustrated in the correlation matrix in the Appendix (Table A.4), younger, smaller, more R&D intensive and technology specialized firms display higher levels of *tech differentiation* relative to older, larger, less R&D intensive and technology diversified firms. Firms with the most unique and differentiated technology portfolio in history across all industries include Pioneer Hi-Bred International in 1997, Monsanto in 2011, Immunomedics in 2015, Olin Corporation in 2015, and Innoviva in 2015.[18] Companies with a persistent high level of technology differentiation in their respective industries over time include Monsanto (agricultural production-crops), Tesla (motor vehicles & passenger car bodies), Alnylam Pharmaceuticals and Celgene (pharmaceutical preparations), Gilead Sciences (biological products), Infinera (computer communications equipment), and Universal Display Corporation and First Solar (semiconductors & related devices).[19] To compare the use of patent text versus patent technology classification or citations to measure the uniqueness and differentiation of firm technology portfolios, we calculate *tech differentiation (class), tech differentiation (subclass),* and *tech differentiation (citation)* in the same way except for using patent classes, subclasses, or backward patent citations instead of keywords.[20] Interestingly, our new *tech differentiation* measure only weakly correlates with *tech differentiation (class)* (corr=0.109), *tech differentiation (subclass)* (corr=0.013), and *tech differentiation (citation)* (corr=-0.074). Tables A.5 and A.6 in the Appendix rank the top 10 firms with the most differentiated technology portfolio in history across all industries and by selected industries and years respectively, and again indicate that the semantic content of patents offers a very different insight in a firm’s competitive position and differentiation in the technology space compared to the traditional empirical approaches. Moreover, in contrast to *tech*

[18] This selection is restricted to firms with at least 100 patents in their portfolio.
[19] Industries defined based on SIC. The selection is restricted to firms with at least 100 patents in their 2015 portfolio.
[20] Table A.1 in the Appendix provides a detailed description on the calculation of each measure.

differentiation based on text, the younger, smaller and more R&D intensive firms do not display higher levels of technology differentiation as measured when based on patent technology classification or citations (see Table A.4 in the Appendix).

2.4 Other variables

To study how technology differentiation relates to firm performance, we use the entire firm-year level panel of U.S. public firms matched to patents, and collect *Tobin's Q*[21] as a proxy for firm performance from Compustat. As a robustness check, we use *ROA* as an alternative measure of firm performance. In line with prior work (e.g. Bellstam et al. 2021, Hirshleifer et al. 2018), we use *total assets*, *leverage*, *cash*, *asset tangibility*, and *firm age* as control variables. In addition, we control for *R&D intensity* (R&D investments/total assets), *citation-weighted patents* (number of patents in the portfolio of firm *i* in year *t* weighted by the number of forward cites received by these patents), and *tech specialization* (the degree to which the patent portfolio of firm *i* in year *t* is concentrated in a small number of technology fields). Moreover, given the important relation between product market competition and firm performance, and particularly between product market competition and firm-level innovation strategy (Aghion et al. 2005), we additionally control for the time varying amount of *prod market competition* faced by the focal firm using data on product market overlap between public firms from Hoberg and Phillips (2016).[22] In line with prior studies, we winsorize all performance and control variables from Compustat (except for *firm age*) at the 1% and 99% levels (e.g. Custódio et al. 2019). We end up with an unbalanced panel of 4,053 firms and 38,550 firm-year level observations for the years 1989-2015. A firm is on average observed over approximately 10 years.[23] All variables are defined in Table A.1 in the Appendix. Table 1 below shows descriptive statistics and Table A.4 in the Appendix displays the correlation matrix.

'Insert Table 1'

2.5 Method

We study how technology differentiation relates to firm performance using the following firm-year level panel model:

---

[21] All findings remain robust if we use an alternative calculation of Tobin's Q which better accounts for intangible assets, amongst others by accounting for all prior R&D investments of the firm (see Peters and Taylor 2017).
[22] Data on product market competition is only available since 1989, so that we lose 4 years of observations in the regression analysis. Nevertheless, all results are robust to including the first 4 years of the panel and excluding product market competition as control variable.
[23] Our results are robust for firm survivor bias, i.e. for the subset of firms which remained active until the last year in our sample (i.e. 2015). Also, all findings remain robust for the subset of firms which are observed over at least 10 years.

$$Y_{it} = \alpha_i + \gamma_j + \delta_t + \beta_1 * tech\ differentiation_{it} + \beta_2 X_{it-1} + \varepsilon_{it} \quad (1)$$

Y_{it} refers to the performance of firm *i* in year *t* as measured by *Tobin's Q* or *ROA*, α_i, γ_j, and δ_t capture firm, industry,[24] and year fixed effects, and X_{it-1} includes all control variables lagged by one year (i.e. *total assets, leverage, cash, asset tangibility, firm age, R&D intensity, citation-weighted patents, tech specialization, prod market competition*). Notice that $tech\ differentiation_{it}$ is measured based on firm *i*'s patents from years *t-5* to *t-1*.[25] Firm and industry fixed effects control for time-invariant unobserved heterogeneity in firm performance across firms and industries, and year fixed effects control for unobserved heterogeneity across years. In Appendix (Tables A.7 and A.8), we illustrate the robustness of our main findings if we additionally control for unobserved heterogeneity at the industry-year and technology(-year) level.[26]

3. Results

As shown in Table 2, *tech differentiation* has a strong positive relation with firm performance significant at the 1% level both in the cross-section and with firm fixed effects. Using firm fixed effects, one standard deviation increase in *tech differentiation* is related to an increase of *Tobin's Q* with 13.4% and *ROA* with 3.3%. The size of the marginal effects of *tech differentiation* are also very significant in comparison with the effects of other traditional firm innovation measures. A one standard deviation increase in *R&D intensity* and *citation-weighted patents* affects *Tobin's Q* with +6.3% and +17.5% and ROA with -8.5%[27] and +1.8% respectively. As such, the differentiation of a technology portfolio arguably seems important for *ROA* and market value relative to the size of the technology portfolio as measured by the number of citation-weighted patents. This is surprising given that *citation-weighted patents* relies on future information in regards to the impact and likely commercial value of individual patents, which only becomes available many years after a patent is granted, while *tech differentiation* exclusively relies on the

[24] Industry fixed effects are generated based on 3-digit SIC codes. We also include industry fixed effects in the models with firm fixed effects because a small number of firms changed SIC codes over time.

[25] Using lagged control variables and *tech differentiation* based on the patent portfolio from years t-1 to t-5 reduces concerns about reverse causality.

[26] Given the number of observations in our sample (especially in the split sample analyses later in the paper), we do not use industry-year and technology(-year) level fixed effects in our main specification because this would include several thousand additional fixed effects besides the several thousand fixed effects (firm-level, industry-level, year-level) already included in our baseline specification. In addition, we do not include technology and technology-year level fixed effects in our main specification because these fixed effects represent an alternative means to characterize the technology portfolio of firms and to map firms' position in technology space based on the classification of patents rather than patent text (e.g. Jaffe, 1986). Including additional fixed effects based on patent classification runs counter our effort to compare the use of patent text versus patent classification to characterize the technology portfolio of firms. Nevertheless, as illustrated in Tables A.7 and A.8 in Appendix, our main effects remain robust after including industry-year level and/or technology and technology-year level fixed effects besides the firm-, industry- and year-level fixed effects.

[27] The negative correlation between the *R&D intensity* and *ROA* of firm *i* in year *t* might reflect that R&D intensity only pays off in terms of ROA in the future while corresponding with higher R&D expenditures and lower *ROA* in the current year. Prior papers found a similar negative correlation between R&D intensity and ROA (e.g. DesJardine and Durand 2020).

technical content of patents at the time they are granted (Hall et al. 2005). In line with prior work in this research stream, we have no exogenous variation and need to be careful in interpreting these results as causal (e.g. Hall et al. 2005, Simeth and Cincera 2016, Bellstam et al. 2021).

‘Insert Table 2’

As illustrated in Appendix, our measure also works for firms with only a few patents in their portfolio as well as for firms with a large patent portfolio (Table A.9). Because a single patent document has on average 61 unique technical keywords, a few patents can provide a good insight in a firm’s technology portfolio. As such, our method could arguably also be used for smaller (non-public) firms and startups. In addition, the method works for firms specialized in a single product market industry as well as for diversified firms operating in multiple industries (Table A.10). Moreover, our findings remain robust if we calculate *tech differentiation* of firm *i* in year *t* exclusively based on *tech similarity* of firm *i* in year *t* with all other firms from the same industry (SIC3) in year *t* (instead of all firms from all industries), or if we calculate *tech differentiation* of firm *i* in year *t* as one minus the maximum (rather than the average) *tech similarity* between firm *i* and all other firms in year *t* (Table A.11). Finally, technology differentiation measured by means of patent technology classes, subclasses, or prior art citations has no statistically significant relation with firm performance, and the size and economic significance of the coefficients is very small compared to *tech differentiation* based on patent text (see columns 3-5 and 8-10 of Table 2). These findings suggest that patent technology classifications or citations do not provide sufficiently granular information to characterize a firm’s technology portfolio and map a firm’s competitive position and differentiation in the technology space.

Among both firm performance measures, the effect of *tech differentiation* on Tobin’s Q is presumably the largest because the market value of a company is also taking into consideration the effect of technology differentiation on future sales and profits. To examine the long-run effect of *tech differentiation* on firm performance, we re-estimate model (1) but vary the timing of the performance measures up to five years in the future. Figure 3 plots the marginal effects of a one standard deviation increase in *tech differentiation* (measured in year t) on *Tobin’s Q* (panel a) and *ROA* (panel b) measured in years t up to t+5; Tables A.12 and A.13 in the Appendix display the corresponding regression tables. Figure 3(a) illustrates that *tech differentiation* has a strong positive relation with *Tobin’s Q*, but the effect gradually declines and becomes insignificant after approximately 5 years. Figure 3(b) shows that the marginal effect of *tech differentiation* on *ROA* peaks in year t, corresponding with a 3.3% increase in *ROA*, followed by a slow decline before becoming statistically insignificant after year t+4.

‘Insert Figure 3’

To better understand the relation between unique and differentiated firm technology and superior firm performance, we split the sample by R&D intensity and product market rivalry in each industry. In R&D intensive industries, firms heavily invest in R&D because pioneering new and differentiated technology is crucial for a firm's competitive advantage and financial performance. Therefore, we expect that the differentiation of the firm's technology portfolio relative to other firms has a stronger effect on firm profitability and market value in industries with high R&D intensity, such as medical equipment and turbine and engines, in comparison to lower R&D intensive industries, such as the food industry and furniture manufacturing. To test this, we collect the R&D intensity of all U.S. public firms from 1988 to 2014 from Compustat, calculate *industry R&D intensity* by taking the average *R&D intensity* of all firms from the same industry, and split the sample by the mean of *industry R&D intensity*. As illustrated in Figure 4,[28] *tech differentiation* has a significantly stronger relation with both *ROA* and *Tobin's Q* for firms in high R&D intensive industries compared to firms in industries with lower R&D intensity. One standard deviation increase in *tech differentiation* corresponds with an increase in *Tobin's Q* of 16% for firms in high R&D intensive industries versus 6% for firms in low R&D intensive industries, and an increase in *ROA* of 4% for firms in high R&D intensive industries versus 2% for firms in low R&D intensive industries.

'Insert Figure 4'

Likewise, in industries with strong product market rivalry such as the computer hardware and fabricated metal products manufacturing industries, firms have a stronger incentive to continuously try to push the technology frontier and escape competition compared to firms in industries with less product market rivalry, such as the beverage, tobacco, and musical instruments industries (Aghion et al. 2005). In case many firms sell similar products to the same customers, firms arguably face a greater risk of losing business to firms with unique and differentiated technology. Therefore, we expect that technology differentiation has a stronger relation with firm profitability and market value for firms in industries with high product market rivalry versus firms in industries with less product market rivalry. To test this, we first calculate *prod market competition* for all U.S. public firms from 1989 to 2015 from Compustat, take the average *prod market competition* of all firms from the same industry as our measure for *industry prod market competition,* and split the sample by the mean of *industry prod market competition*. Figure 5[29] shows that *tech differentiation* has a significantly stronger relation with *Tobin's Q* and *ROA* for firms in industries with higher product market rivalry. One standard deviation increase in *tech differentiation* relates to an increase in *Tobin's Q* of 18% for firms in industries with high product market rivalry versus

[28] Table A.14 (columns 1-4) in Appendix shows the corresponding regressions.
[29] Table A.14 (columns 5-8) in Appendix shows the corresponding regressions.

4% for firms in industries with less product market rivalry, and an increase in *ROA* of 5% for firms in more competitive product market industries versus 1% for firms in industries with less product market rivalry. As a robustness check, we use two alternative measures for the product market rivalry of an industry. First, we calculate a Herfindahl-index (HHI) for industry concentration using the sales of all public firms in Compustat from 1988-2014. Second, we collect the HHI for the concentration of each industry from the 2017 U.S. Census, which includes both public and private firms. In line with Figure 5, we find that *tech differentiation* has a significantly stronger relation with *Tobin's Q* and *ROA* for firms in industries with high product market rivalry versus firms in industries with low product market rivalry.[30] As the only exception, *tech differentiation* does not have a significantly stronger relationship with *ROA* for firms in industries with stronger product market rivalry as measured using U.S. Census-based HHI.

'Insert Figure 5'

4. Discussion and conclusion

Whereas the economics and strategy literature has mainly focused on a firm's competitive position and differentiation in the product market, we introduce and empirically demonstrate the importance of a firm's competitive position and differentiation in the technology space. Using a panel of all U.S. public firms matched to patents for the years 1980-2015, our findings suggest that a firm's competitive advantage and superior performance relies on a unique and differentiated technology portfolio, particularly in R&D intensive industries and in industries with strong product market rivalry. To do so, we develop a new method to characterize the technology portfolio of firms based on the semantic content of patents, to map each firm's competitive position in technology space relative to all other firms, and to measure the uniqueness or differentiation of a firm's technology portfolio. We show that our method provides a very different characterization of firm technology portfolios compared to the conventional approach based on patent technology classification or citations and more strongly correlates with firm profitability and market value. Finally, we provide open access to all code and data to measure the technology similarity between and the technology differentiation of U.S. public firms between 1980 and 2015. While we do not uncover the particular theoretical mechanisms relating technology differentiation to firm performance, we hope that our work and data unlock many different avenues for future research in this exciting area.

[30] Table A.14 (columns 9-16) in Appendix shows the corresponding regressions.

5. References

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-U relationship. *The Quarterly Journal of Economics*, 120(2), 701-728.

Ahuja, G., Katila, R., (2001). Technological Acquisitions and the Innovation Performance of Acquiring Firms: A Longitudinal Study. *Strategic Management Journal*, 22,197-220.

Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic management journal*, 22(6-7), 521-543.

Arora, A., Gambardella, A., (1990). Complementarity and External Linkages: The Strategies of The Large Firms in Biotechnology. *The Journal of Industrial Economics*, 38(4), 361-379.

Arora, A., Fosfuri, A., Gambardella, A., (2001). Markets for Technology and Their Implications for Corporate Strategy. *Industrial and Corporate Change*, 10(2),419-452.

Arora, A., Belenzon, S., & Sheer, L. (2021). Matching patents to Compustat firms, 1980–2015: Dynamic reassignment, name changes, and ownership structures. *Research Policy*, 50(5), 104217.

Arts, S., Cassiman, B., & Gomez, J.C., (2018). Text Matching to Measure Patent Similarity. *Strategic Management Journal*, 39(1), 62-84.

Arts, S., Hou, J., & Gomez, J.C. (2021a). Natural Language Processing to Identify the Creation and Impact of New Technologies in Patent Text: Code, Data, and New Measures. *Research Policy*, 50(2), 104-144.

Arts, S., Cassiman, B., & Hou, J. (2021b). Technology differentiation, product market rivalry, and M&A transactions. *Working paper*.

Bellstam, G., Bhagat, S., & Cookson J. A., (2021). A Text-Based Analysis of Corporate Innovation. *Management Science*, 67(7): 4004-4031.

Bloom, N., Schankerman, M., Van Reenen, J., (2013). Identifying Technology Spillovers and Product Market Rivalry. *Econometrica*, 81(4), 1347-1393.

Custódio, C., Ferreira, M. A., & Matos, P., (2019). Do General Managerial Skills Spur Innovation? *Management Science*, 65(2), 459-476.

DesJardine, M. R., & Durand, R. (2020). Disentangling the effects of hedge fund activism on firm financial and social performance. *Strategic Management Journal*, 41(6), 1054-1082.

Gil Moltó, MJ., Georgantzís, N. and V. Orts. (2005). Cooperative R&D with Endogenous Technology Differentiation. *Journal of Economics and Management Strategy*, 14(2), 461–476.

Guzman, J., & Li, A. (2019). Measuring Founding Strategy. Forthcoming *Management Science*.

Hall, B. H., Jaffe, A.B., & Trajtenberg, M., (2005). Market Value and Patent Citations. *RAND Journal of Economics*, 36(1), 16-38.

Hall, B. H., Helmers, C., Rogers, M., & Sena, V., (2014). The Choice between Formal and Informal Intellectual Property: A Review. *Journal of Economic Literature*, 52(2), 375-423.

Hirshleifer, D., Hsu, P., & Li, D., (2018). Innovative Originality, Profitability, and Stock Return. *The Review of Financial Studies*, 31(7), 2553-2605.

Hoberg, G., Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, 124(5), 1423-1465.

Jaffe, A. B., (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. *The American Economic Review*, 76(5), 984-1001.

Jaffe, A.B., (1989). Characterizing the “Technological Position” of Firms, with Application to Quantifying Technological Opportunity and Research Spillovers. *Research Policy*, 18(2), 87-97.

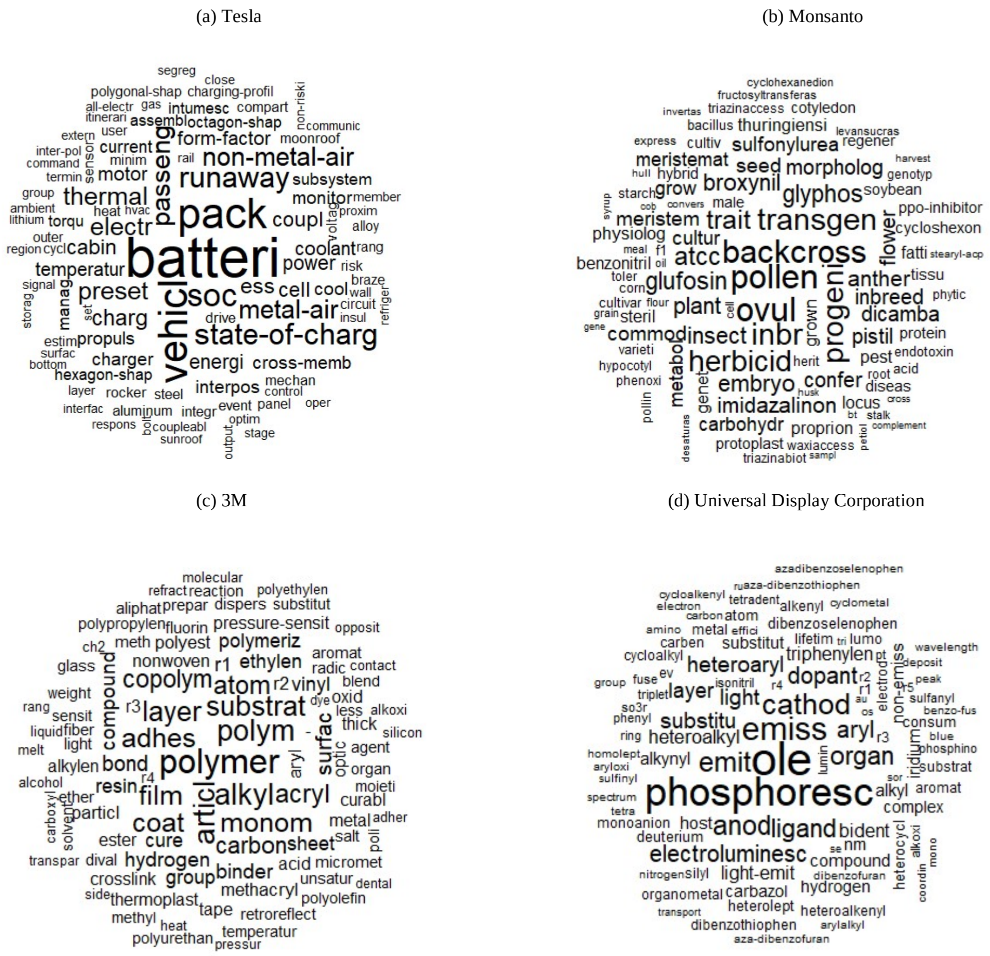
Kamien, M.I. and Zang, I., (2000). Meet me halfway: Research Joint Ventures and Absorptive Capacity. *International Journal of Industrial Organization*, 18, 995–1012.

Lin, P. and Zhou, W., (2013). The Effects of Competition on the R&D Portfolios of Multiproduct firms. *International Journal of Industrial Organization*, 31, 83–91.

Mowery, D.C., Oxley, J.E. & Silverman, B.S., (1998). Technological overlap and interfirm cooperation: implications for the resource-based view of the firm. Research Policy, 27, 507–523.Peteraf, M. A.,

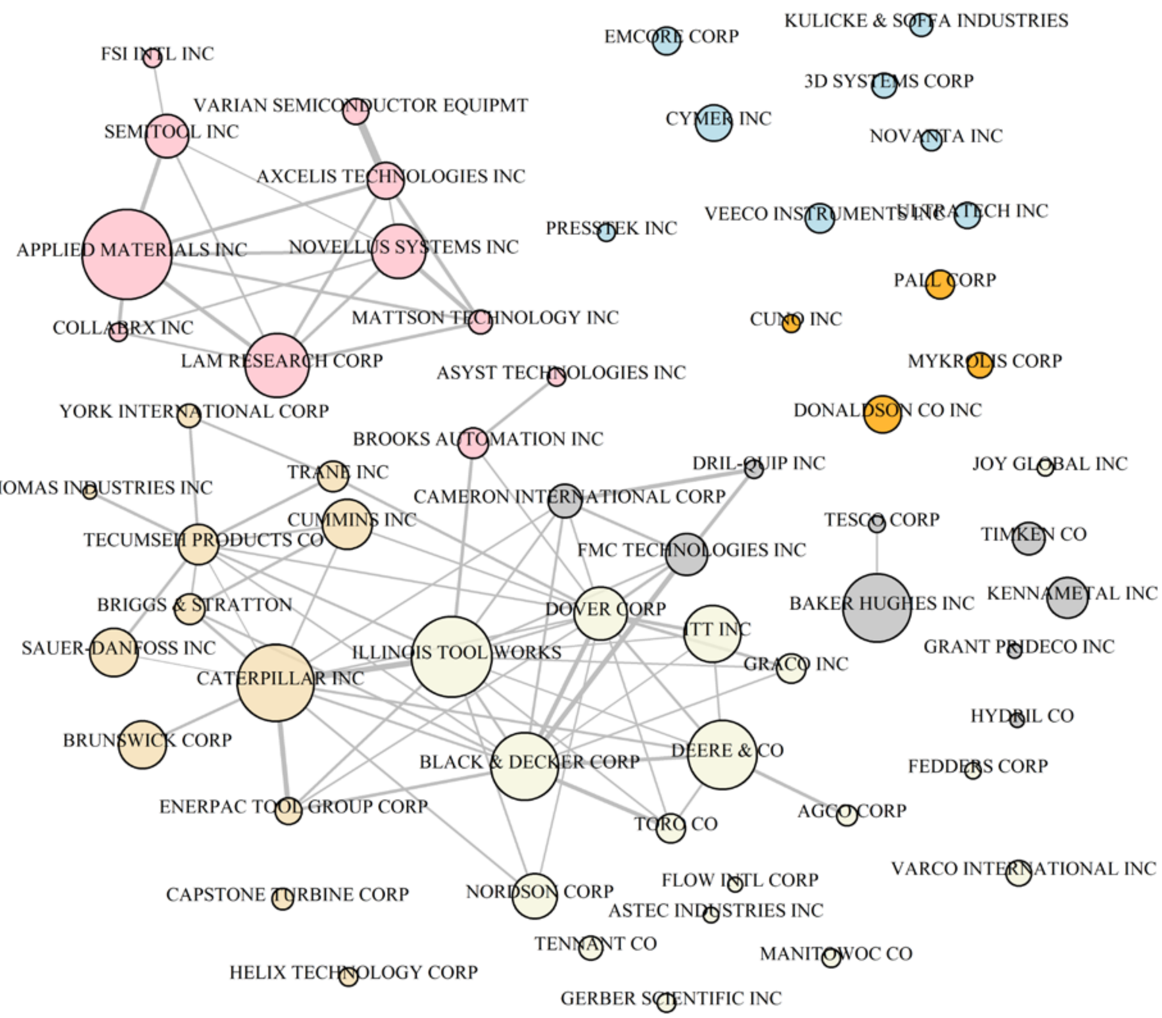
(1993). The Cornerstones of Competitive Advantage: A Resource-Based View. *Strategic Management Journal*, 14(3), 179-191.
Peters, R. H., & Taylor, L. A. (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics*, 123(2), 251-272.
Righi, C., Simcoe, T. (2019). Patent Examiner Specialization. *Research Policy*, 48(1), 137-148.
Rosenkopf, L., Almeida, P., (2003). Overcoming Local Search through Alliances and Mobility. *Management Science*, 49(6), 751-766.
Rothaermel, F. T., Deeds, D. L., (2004). Exploration and Exploitation Alliances in Biotechnology: A System of New Product Development. *Strategic Management Journal*, 25(3), 201-221.
Silverman, B.S., (1999) Technological Resources and the Direction of Corporate Diversification: Toward an Integration of the Resource-Based View and Transaction Cost Economics. *Management Science*, 45(8), 1109-1124.
Simeth, M., Cincera, M., (2016). Corporate Science, Innovation, and Firm Value. *Management Science*, 62(7), 1970-1981.
Stuart, T. E., & Podolny, J. M. (1996). Local Search and the Evolution of Technological Capabilities. *Strategic Management Journal*, 17(S1), 21-38.
Thompson, P., Fox-Kean M. (2005). Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. *American Economic Review*: 450-460.
Wernerfelt, B., (1984). A Resource-Based View of the Firm. *Strategic Management Journal*, 5(2), 171-180.

Figure 1. Characterizing firm technology portfolios using word clouds



Notes: This figure illustrates a simplified version of our characterization of firm technology portfolios using word clouds for four companies, namely Tesla (panel a), Monsanto (panel b), 3M (panel c), and Universal Display (panel d). For each firm, we plot its top 100 most frequently used stemmed technical keywords identified from its patents granted between 1980 and 2015, word size is proportional to TF-IDF weights.

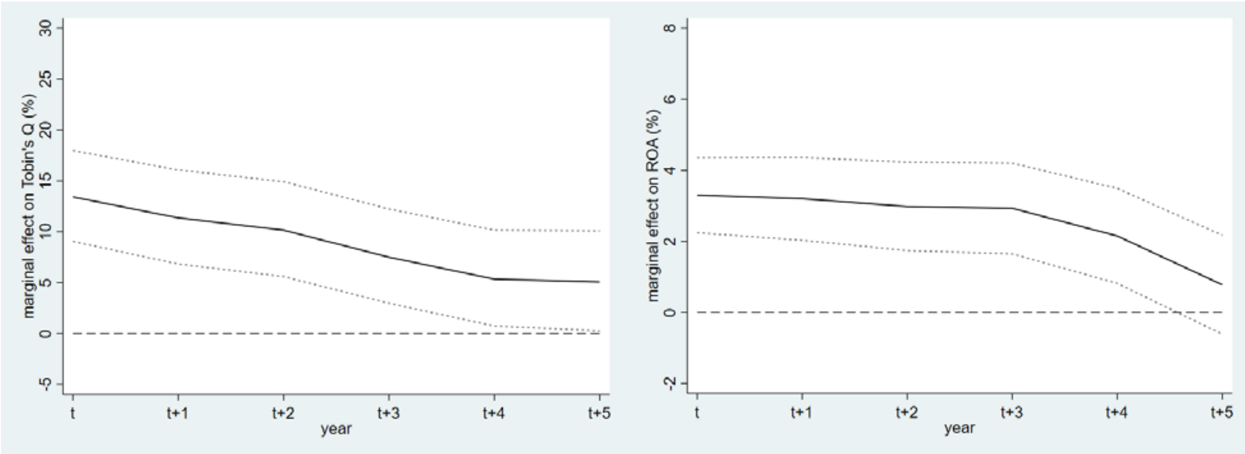
Figure 2. Network of firms in machinery industry in 2005



Notes: This network graph displays all firms from the machinery industry (as defined by Fama and French, 1997) in 2005 with at least 20 granted patents filed between 2000 and 2004. Nodes represent firms, node size is proportional to the number of granted patents filed by the firm between 1995 and 1999. Two nodes are connected by an edge if *tech similarity* is larger than 0.6, and edge thickness is proportional to *tech similarity* between firms (thicker edge means more similar firms). Node colors indicate clusters of similar firms based on hierarchical clustering using the Ward method. Six clusters are identified, but this number is not determined by the algorithm and can change according to personal preferences. The pink cluster largely corresponds to semiconductor equipment technology firms (e.g. Applied Materials), the dark brown cluster to engine technology firms (e.g. Caterpillar), the orange cluster to fluid and air filter technology firms (e.g. Donaldson Company), the grey cluster to mining, oil, and gas drilling technology firms (e.g. Baker Hughes), the blue cluster to photonics and laser technology firms (e.g. Veeco Instruments), and the light brown cluster to power tools and component technology firms (e.g. Black & Decker).

Figure 3. Long term effect of technology differentiation on firm performance

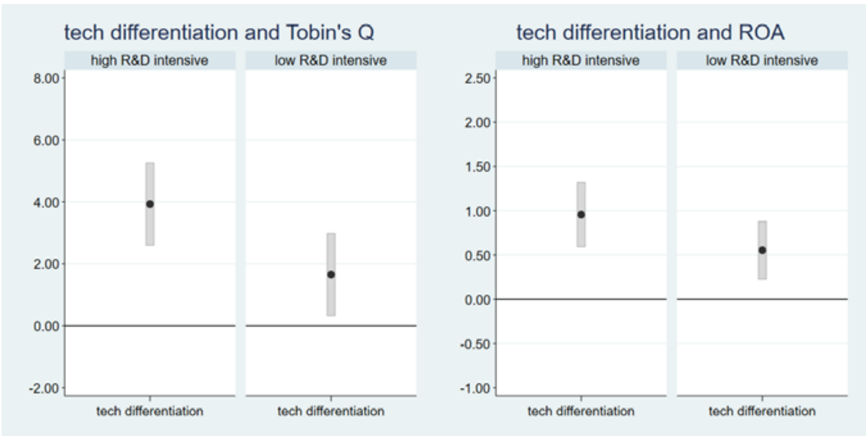
(a) Tobin's Q (b) ROA



Notes: The graphs illustrate the marginal effects (in %) and the corresponding 95% confidence intervals of a one standard deviation increase in *tech differentiation* in year *t* on *Tobin's Q* and *ROA* measured in years *t* to *t+5*. Regression results can be found in Tables A.12 and A.13 in Appendix.

Figure 4. Technology differentiation and firm performance for high versus low R&D intensive industries

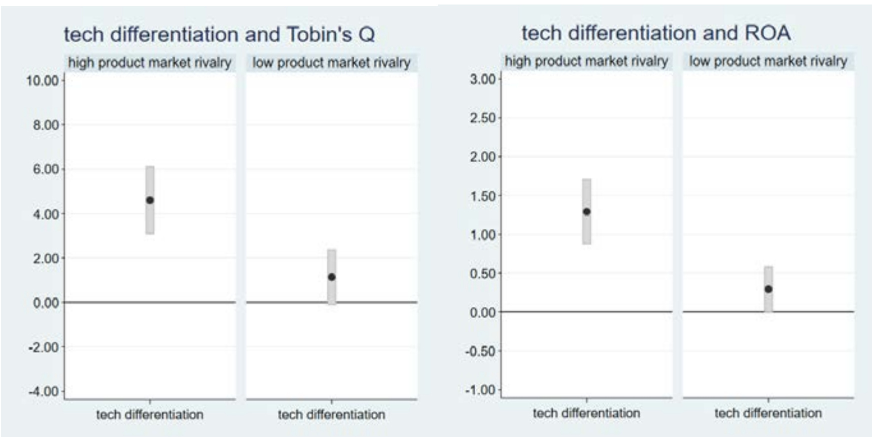
(a) Tobin's Q (b) ROA



Notes: The graphs illustrate estimated coefficients and the corresponding 95% confidence intervals of *tech differentiation* on *Tobin's Q* and *ROA* for samples split by the mean of *industry R&D intensity*. Regression results can be found in Table A.14 in Appendix.

Figure 5. Technology differentiation and firm performance for industries with high versus low product market rivalry

(a) Tobin's Q (b) ROA



Notes: The graphs illustrate estimated coefficients and the corresponding 95% confidence intervals of *tech differentiation* on *Tobin's Q* and *ROA* for samples split by the mean of *industry prod market competition*. Regression results can be found in Table A.14 in Appendix

Table 1: Descriptive statistics

		Mean	Median	S.D.	Min	Max
(1)	Tobin's Q	2.285	1.532	2.506	0.115	22.612
(2)	ROA	-2.922	9.315	39.341	-292.807	37.272
(3)	Tech differentiation	0.940	0.949	0.036	0.783	0.998
(4)	Tech differentiation (class)	0.956	0.963	0.043	0.000	1.000
(5)	Tech differentiation (subclass)	0.995	0.998	0.034	0.000	1.000
(6)	Tech differentiation (citation)	0.995	1.000	0.068	0.000	1.000
(7)	R&D intensity	12.235	6.213	18.493	0.000	133.681
(8)	Citation-weighted patents	837.518	59.000	4,526.416	1.000	170,542.000
(9)	Tech specialization	0.452	0.360	0.319	0.000	1.000
(10)	Prod market competition	13.628	5.728	17.734	0.000	100.118
(11)	Total assets	2,175.337	172.476	6,436.666	0.618	41,575.000
(12)	Leverage	17.892	11.872	21.771	0.000	178.273
(13)	Cash	27.891	18.545	26.875	0.077	95.026
(14)	Asset tangibility	19.850	15.791	15.884	0.208	72.833
(15)	Firm age	13.869	13.000	9.072	0.000	39.000

Notes: This table reports the descriptive statistics of the sample used to examine the relationship between tech differentiation and firm performance from 1989 to 2015, and includes 38,550 firm-year observations and 4,053 firms. *ROA*, *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* are measured as percentages. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. All financial measures from Compustat are winsorized at levels of 1% and 99%. Definitions of variables can be found from Table A.1 in Appendix.

Table 2: Technology differentiation and firm performance

	Tobin's Q					ROA				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Tech differentiation	1.951***	3.469***				1.082***	0.908***			
	(0.416)	(0.552)				(0.146)	(0.149)			
Tech differentiation (class)			0.119					0.027		
			(0.151)					(0.041)		
Tech differentiation (subclass)				-0.031					-0.015	
				(0.111)					(0.030)	
Tech differentiation (citation)					-0.025					0.017
					(0.082)					(0.036)
R&D intensity	0.837***	0.329***	0.315***	0.315***	0.315***	-0.912***	-0.459***	-0.462***	-0.462***	-0.462***
	(0.057)	(0.059)	(0.059)	(0.059)	(0.059)	(0.033)	(0.037)	(0.037)	(0.037)	(0.037)
Citation-weighted patents	0.106***	0.079***	0.050***	0.050***	0.050***	0.005**	0.009***	0.001	0.001	0.001
	(0.007)	(0.009)	(0.008)	(0.008)	(0.008)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Firm fixed effects	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	38,247	38,247	38,247	38,247	38,247	38,375	38,375	38,375	38,375	38,375
Number of firms	4,049	4,049	4,049	4,049	4,049	4,049	4,049	4,049	4,049	4,049
Within r2		0.159	0.156	0.156	0.156		0.076	0.075	0.075	0.075
Between r2		0.076	0.074	0.074	0.074		0.407	0.411	0.411	0.411
Overall r2	0.191	0.068	0.065	0.065	0.065	0.466	0.318	0.326	0.325	0.325
	Marginal effects (in %)									
Tech differentiation	7.34	13.42				3.93	3.30			
Tech differentiation (class)			0.52					0.12		
Tech differentiation (subclass)				-0.10					-0.05	
Tech differentiation (citation)					-0.17					0.12
R&D intensity	16.68	6.25	5.98	5.97	5.97	-16.85	-8.47	-8.54	-8.54	-8.54
Citation weighted patents	24.21	17.45	10.67	10.69	10.71	1.08	1.77	0.22	0.22	0.20

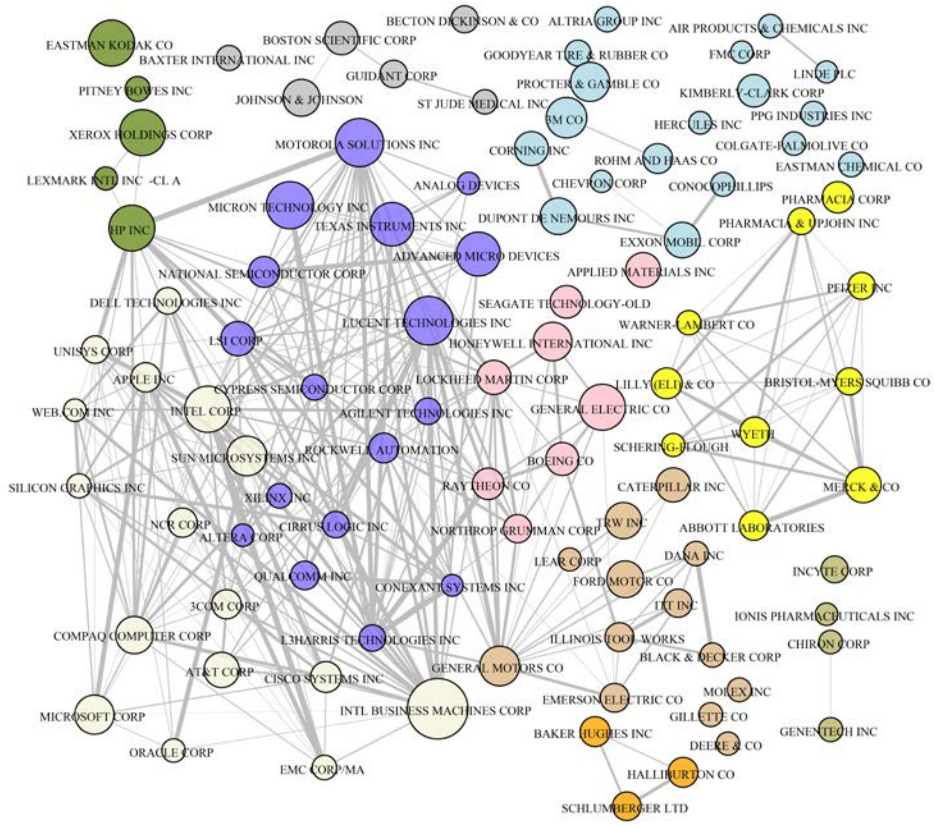
Notes: The table reports coefficient estimates from a linear firm-fixed effects regression (except columns (1) and (6) that are standard OLS regressions). The sample is an unbalanced panel with firm fiscal years ranging from 1989 to 2015. As a result of missing values of firm performance indicators in some years, the number of observations varies across columns. *Tobin's Q* and *citation-weighted patents* are log transformed. Additional control variables include *Total assets (log), Firm age (log), Leverage, Cash, Asset tangibility, tech specialization,* and *prod market competition (log).* We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. Control variables are lagged by one year. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on 3-digit SIC. Definitions of variables are provided in Table A.1 in Appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable caused by a one-standard deviation increase of the corresponding explanatory variable. *, **, and *** denotes significance at the 10%, 5%, and 1% level.

Online Appendix

Figure A.1: Network of firms in machinery industry in 2005 based on tech similarity (class)

Notes: This network graph displays all firms from the machinery industry (as defined by Fama and French, 1997) in 2005 with at least 20 granted patents filed between 2000 and 2004. Nodes represent firms, node size is proportional to the number of granted patents filed by the firm between 1995 and 1999, two nodes are connected by an edge if *tech similarity (class)* is larger than 0.6, and edge thickness is proportional to *tech similarity (class)* between firms (thicker edge means more similar firms). Node colors indicate clusters of similar firms based on hierarchical clustering using the Ward method. Compared to Figure 2, firms are less connected and grouped into different clusters. Fluid and air filter technology firms (e.g. Donaldson) merge with power tools and component technology firms, and photonics and laser technology firms (e.g. Veeco) merge with semiconductor technology firms.

Figure A.2. Network top 100 patenting firms in 2000



Notes: The graph illustrates the network of the top 100 firms in 2000 with the largest number of granted patents filed between 1995 and 1999. Nodes represent firms, node size is proportional to the number of patents filed by the firm between 1995 and 1999 (larger node means larger patent portfolio), two nodes are connected by an edge if *tech similarity* is larger than 0.65, and edge thickness is proportional to *tech similarity* between the firms (thicker edge means more similar firms). Node colors indicate clusters of similar firms based on hierarchical clustering using the Ward method. Ten clusters are identified, but this number is not determined by the algorithm and can change according to personal preferences. The yellow cluster largely corresponds to pharmaceutical technology firms (e.g. Pfizer), the grey cluster to medical equipment technology firms (e.g. Guidant), the light green cluster to biotechnology firms (e.g. Genentech), the dark green cluster to printing technology firms (e.g. Xerox), the purple cluster to semiconductor technology firms (e.g. Texas Instruments), the blue cluster to chemical and consumer goods technology firms (e.g. Du Pont De Nemours), the light brown cluster to computer hardware and software technology firms (e.g. IBM), the brown cluster to car and electric technology firms (e.g. General Motors), the orange cluster to oil field service technology (e.g. Baker Hughes), and the pink cluster to aerospace and defense technology firms (e.g. Lockheed Martin).

Table A.1: Definitions of Variables

Measure	Description
Tech similarity	First, we construct the patent portfolio for firm *i* in year *t* by collecting all granted patents linked to firm *i* which were filed by between year *t-5* and *t-1*. Second, the patent portfolio of firm *i* in year *t* is represented as a vector $S_{it} = (S_{it1}, S_{it2}, \ldots, S_{itK})$, where $k \in (1,K)$ indicates one stemmed technical keyword identified from the entire patent vocabulary and S_{itk} denotes the share of patents from the patent portfolio of firm *i* in year *t* using the given word *k*. Tech similarity between firm *i* and *j* in year *t* is calculated as the cosine between the two vectors ($cos(S_{it}, S_{jt})$) and uses TF-IDF weights.
Tech differentiation	*Tech differentiation* of firm *i* in year *t* is calculated as $1 - \frac{1}{n-1}\sum_{j=1,j\neq i}^{n-1} tech\ similarity_{ijt}$, with *n* equal to all firms active in year *t* and $tech\ similarity_{ijt}$ equal to the technology similarity between firm *i* and firm *j* in year *t*.
Tech differentiation (class)	Calculated in the same way as *tech differentiation* except for using main patents classes instead of keywords. First, the patent portfolio of firm *i* in year *t* is represented as a vector $S_{it} = (S_{it1}, S_{it2}, \ldots, S_{itK})$, where $k \in (1,K)$ indicates one main patent class from the US patent classification system and S_{itk} denotes the share of patents from the patent portfolio of firm *i* in year *t* in patent class *k*. Next, we use these firm-year level vectors and follow the same steps as before to calculate the technology similarity for all pairs of firms for all years and the technology differentiation of each firm for each year.
Tech differentiation (subclass)	Calculated in the same way as *tech differentiation* except for using patent subclasses instead of keywords. First, the patent portfolio of firm *i* in year *t* is represented as a vector $S_{it} = (S_{it1}, S_{it2}, \ldots, S_{itK})$, where $k \in (1,K)$ indicates one patent subclass from the US patent classification system and S_{itk} denotes the share of patents from the patent portfolio of firm *i* in year *t* in patent subclass *k*.. Next, we use these firm-year level vectors and follow the same steps as before to calculate the technology similarity for all pairs of firms for all years and the technology differentiation of each firm for each year.
Tech differentiation (citation)	Calculated in the same way as *tech differentiation* except for using backward patent (prior art) citations instead of keywords. First, the patent portfolio of firm *i* in year *t* is represented as a vector $S_{it} = (S_{it1}, S_{it2}, \ldots, S_{itK})$, where $k \in (1,K)$ indicates one cited patent and S_{itk} denotes the share of patents from the patent portfolio of firm *i* in year *t* which cite patent *k*. Next, we use these firm-year level vectors and follow the same steps as before to calculate the technology similarity for all pairs of firms for all years and the technology differentiation of each firm for each year.
R&D intensity	Research and development investments scaled by total assets.
Citation-weighted patents	The number of patents in the patent portfolio of firm *i* in year *t*, i.e. all granted patents linked to firm *i* which were filed by between year *t-5* and *t-1*, weighted by the number of citations received by these patents within 5 year after grant.
Tech specialization	First, the patent portfolio of firm *i* in year *t* is represented as a vector $S_{it} = (S_{it1}, S_{it2}, \ldots, S_{itK})$, where $k \in (1,K)$ indicates one main patent class and S_{itk} denotes the share of patents from the patent portfolio of firm *i* in year *t* in patent class *k*. Next, firm *i*'s tech specialization in year *t* is calculated as a Herfindahl index based on the share of patents in each class.
Tobin's Q	The ratio of the market value of a firm and the replacement (book) value of the firm's assets. A firm's market value is defined as the sum of market capitalization (share price multiplied by the number of common shares outstanding at the end of the year), preferred stock, minority interests, and total debt minus cash.
ROA	Earnings before interest, taxes, depreciation, and amortization (EBITDA) scaled by total assets.
Total assets	Total assets (in million USD).
Leverage	Total debt (long-term debt and debt in current liabilities) scaled by total assets.
Cash	Cash and short-term investment scaled by total assets.
Asset tangibility	Property plant and equipment divided by total assets.
Firm age	The number of years since the firm first entered Compustat (earliest date 1975).
Prod market competition	In line with Hoberg and Phillips (2016), we take the sum of pairwise product similarity scores (based on the business descriptions from annual 10-K reports) between the focal firm *i* in year *t* and all other US public firms from the same 3-digit SIC industry in year *t* (regardless of firm's patenting activities). Data available from https://hobergphillips.tuck.dartmouth.edu/tnic_poweruser.htm.
Industry R&D intensity	First, we collect the R&D intensity of all firms from 1988 to 2014 from Compustat. Second, we take the average of R&D intensity of all firms from the same 3-digit SIC industry across years.
Industry prod market competition	The average product market competition of all firms from the same 3-digit SIC industry across years.
Industry prod market competition (Compustat-based HHI)	First, we collect sales of all firms from 1988 to 2014 from Compustat. Second, for each 3-digit SIC industry-year, we calculate a Herfindahl index (HHI) of industry concentration using the market shares of all firms from the same industry. Finally, we take the average HHI across years for each 3-digit SIC industry.
Industry prod market competition (U.S. Census-based HHI)	The U.S. Census Bureau directly provides the HHI of industry concentration based on the sales of the largest 50 firms within each 3-digit NAICS industry in 2017. We download the data from United States Census Bureau: https://data.census.gov/cedsci/table?q=concentration&y=2017&n=N0300.00&tid=ECNSIZE2017.EC1700SIZECONCEN&hidePreview=true.

Table A.2: Top 10 most similar firms by tech similarity

Focal firm	Rank	Text	Classes	Subclasses	Citations
	1	Ford Motor	Ford Motor	Ford Motor	Boeing
General Motors (2008)	2	Caterpillar	Caterpillar	Cummins	Ford Motor
	3	Cummins	Deere & Co	Caterpillar	Caterpillar
	4	General Electric	Telenav	Tenneco	Borgwarner
	5	Deere & Co	SPX	Tesla	Cummins
	6	Borgwarner	Cummins	Deere & Co	Immersion
	7	Emerson Electric	Tesla	Borgwarner	Raytheon
	8	Honeywell International	Rockwell Collins	Boeing	PPG Industries
	9	Visteon	Borgwarner	TRW Automotive	Apple
	10	Parker-Hannifin	Honeywell International	Visteon	Synaptics
	1	3M	3M	Honeywell International	Air Products & Chemicals
Du Pont De Nemours (1994)	2	Rhone-Poulenc Rorer	Grace (W R) & Co	3M	Honeywell International
	3	Rohm & Haas	Rohm & Haas	Occidental Petroleum	AMP
	4	PPG Industries	PPG Industries	Rhone-Poulenc Rorer	3M
	5	Goodrich	Hercules	Amoco	Eastman Kodak
	6	Union Carbide	Eastman Chemical	Pharmacia & Upjohn	Grace (W R) & Co
	7	Exxon Mobil	Goodrich	PPG Industries	PPG Industries
	8	Grace (W R) & Co	Corning	Rohm & Haas	Molex
	9	Honeywell International	Aerojet Rocketdyne	Eastman Kodak	Pharmacia & Upjohn
	10	Eastman Kodak	Union Carbide	Grace (W R) & Co	Bristol-Myers Squibb
	1	Advanced Micro Devices	Advanced Micro Devices	Texas Instruments	Applied Materials
Micron Technology (2003)	2	Texas Instruments	Amkor Technology	Advanced Micro Devices	Web.com
	3	Agere Systems	Xperi Holding	LSI	Integrated Device Tech
	4	National Semiconductor	Fairchild Semiconductor INTL	Agere Systems	Advanced Micro Devices
	5	LSI	Cree	National Semiconductor	INTL Business Machines
	6	Fairchild Semiconductor INTL	Atmel	Amkor Technology	Texas Instruments
	7	Applied Materials	INTL Rectifier	Cypress Semiconductor	LSI
	8	Cypress Semiconductor	Texas Instruments	Xperi Holding	Novellus Systems
	9	Atmel	National Semiconductor	Intel	Amkor Technology
	10	On Semiconductor	Applied Materials	Applied Materials	Lam Research
	1	INTL Business Machines	L3harries Technologies	INTL Business Machines	INTL Business Machines
AT&T (1986)	2	Rockwell Automation	Rockwell Automation	Motorola Solutions	Motorola Solutions
	3	Honeywell	INTL Business Machines	North American Philips	ITT
	4	L3harries Technologies	ITT	Texas Instruments	Unisys
	5	North American Philips	GTE	L3harries Technologies	Texas Instruments
	6	Texas Instruments	Texas Instruments	RCA	Data General
	7	Motorola Solutions	National Semiconductor	Advanced Micro Devices	North American Philips
	8	ITT	Motorola Solutions	ITT	NCR
	9	HP	Fairchild Semiconductor INTL	National Semiconductor	GTE
	10	GTE	Advanced Micro Devices	GTE	Honeywell International
	1	Chevron	Shell Oil	Chevron	Occidental Petroleum
Gulf (1983)	2	Union Carbide	Chevron	Exxon Mobil	Chevron
	3	Exxon Mobil	Exxon Mobil	Occidental Petroleum	Newmarket
	4	Du Pont	Conocophilips	Union Carbide	Union Carbide
	5	Shell Oil	Ashland Global Holdings	Conocophilips	Exxon Mobil
	6	Conocophilips	Atlantic Richfield	Shell Oil	Conocophilips
	7	Texaco	Union Carbide	Celanese	Du Pont
	8	Atlantic Richfield	Pharmacia	Du Pont	AMF
	9	Pharmacia	Quantum Chemical	Pharmacia	Texaco
	10	Newmarket	Texaco	Navistar International	Goodrich

Notes: The table lists for each focal firm the top 10 most similar firms (ranked by similarity to focal firm, most similar firm in top row) based on *tech similarity*, *tech similarity (class)*, *tech similarity (subclass)*, and *tech similarity (citation)* respectively. We select a number of well-known firms with the largest patent portfolios in their industry-year cohorts as focal firms (industry defined by 2-digit SIC). When selecting the most similar firms, we restrict the sample to firms with at least 100 patents in the portfolio. The average overlap between the top 10 most similar firms identified by *tech similarity* and those identified by *tech similarity (class)*, *tech similarity (subclass)*, and *tech similarity (citation)* is 0.54, 0.62, and 0.50 respectively.

Table A.3: Variance decomposition of technology differentiation

	Decomposition	
	Variance	% of total variance
Total	50.747	100
Between different industries	14.027	27.64
Between different firms in the same industry	31.141	61.37
Within the same firm across years	6.003	11.83

Notes: The table illustrates the variance of *tech differentiation* from different sources. We denote the *tech differentiation* of firm *i* from industry *j* in year *t* as $tech\ diff_{ijt}$. Then total variance is calculated as $\sum_i \sum_j \sum_t \left(tech\ diff_{ijt} - \overline{\overline{\overline{tech\ diff}}}\right)^2$, where $\overline{\overline{\overline{tech\ diff}}}$ represents the mean of tech differentiation based on the whole sample. The decomposition scheme decomposes the variance of tech differentiation into variance across industries $\sum_i \sum_j \sum_t \left(\overline{\overline{tech\ diff_j}} - \overline{\overline{\overline{tech\ diff}}}\right)^2$, variance within a firm over years $\sum_i \sum_j \sum_t \left(tech\ diff_{ijt} - \overline{tech\ diff_{ij}}\right)^2$, and variance across firms within an industry $\sum_i \sum_j \sum_t \left(\overline{tech\ diff_{ij}} - \overline{\overline{tech\ diff_j}}\right)^2$, where $\overline{tech\ diff_{ij}}$ and $\overline{\overline{tech\ diff_j}}$ represents the mean of *tech differentiation* within firm *i* and the mean of *tech differentiation* within industry *j* respectively.

Table A.4: Correlation matrix

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1)	Tobin's Q	1.00														
(2)	ROA	-0.36	1.00													
(3)	Tech differentiation	0.14	-0.28	1.00												
(4)	Tech differentiation (class)	-0.09	0.13	0.07	1.00											
(5)	Tech differentiation (subclass)		0.01	0.03	0.81	1.00										
(6)	Tech differentiation (citation)		0.02	-0.06	0.37	0.49	1.00									
(7)	R&D intensity	0.36	-0.60	0.23	-0.22	-0.03		1.00								
(8)	Citation-weighted patents		0.08	-0.37	-0.12	-0.06		-0.04	1.00							
(9)	Tech specialization	0.08	-0.18	0.65	0.23	0.10	-0.06	0.11	-0.18	1.00						
(10)	Prod market competition	0.20	-0.32	0.27	-0.33	-0.06		0.47	-0.01	0.14	1.00					
(11)	Total assets	-0.06	0.15	-0.44	-0.14	-0.05	0.01	-0.15	0.51	-0.29	-0.10	1.00				
(12)	Leverage		0.03	-0.07	0.04	0.02	0.02	-0.06	0.02	-0.08	-0.15	0.11	1.00			
(13)	Cash	0.25	-0.42	0.28	-0.24	-0.05	-0.02	0.46	-0.05	0.17	0.57	-0.19	-0.34	1.00		
(14)	Asset tangibility	-0.10	0.20	-0.08	0.12	0.04	0.03	-0.19	0.02	-0.14	-0.33	0.15	0.27	-0.50	1.00	
(15)	Firm age	-0.16	0.26	-0.32	0.03			-0.26	0.15	-0.25	-0.30	0.35	0.13	-0.36	0.15	1.00

Notes: This table reports the correlation between variables based on the sample used to examine the relationship between tech differentiation and firm performance from 1989 to 2015, and includes 38,550 firm-year observations and 4,053 firms. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. All financial measures from Compustat are winsorized at levels of 1% and 99%. Definitions of variables can be found from Table A.1 in Appendix. Only correlation coefficients significant at 0.05 are displayed.

Table A.5: Top 10 firms with the most differentiated technology portfolio in history

Text			Classes		
	year	tech differentiation		year	tech differentiation (class)
Monsanto	2011	0.984	Callaway Golf	2007	0.996
Immunomedics	2015	0.984	Resmed	2015	0.993
Pioneer Hi-bred International	1997	0.982	Kennametal	2011	0.992
Olin	2015	0.981	Mattel	2015	0.992
Pharmacia & Upjohn	1981	0.978	Smith International	2001	0.991
Lexicon Pharmaceuticals	2005	0.977	INTL Game Technology	2015	0.991
Innoviva	2015	0.975	Beam	2008	0.991
Mycogen	1996	0.974	Newmarket	1998	0.990
Inois Pharmaceuticals	2015	0.973	Align Technology	2006	0.990
Neurogen	2000	0.972	Lubrizol	1998	0.989
Subclasses			Citations		
	year	tech differentiation (subclass)		year	tech differentiation (citation)
Callaway Golf	2007	1.000	Mattel	1987	1.000
Zygo	2008	1.000	Sunedison	2005	1.000
Albany International	2007	1.000	Innoviva	2015	1.000
K2	2007	0.999	International Flavors & Fragrances	1980	1.000
Dexcom	2011	0.999	K2	2007	1.000
Resmed	2015	0.999	Wabtec	2006	1.000
Timken	2004	0.999	Celanese	2015	1.000
Mattel	1987	0.999	Cymer	2000	1.000
Graphic Packaging	2000	0.999	Masco	2001	1.000
Hayes Lemmerz International	2003	0.999	Remy International	2015	1.000

Notes: The table ranks the top 10 firms with the most unique and differentiated technology portfolio in history as measured by *tech differentiation, tech differentiation (class), tech differentiation (subclass)*, and *tech differentiation (citation)* respectively. The selection is restricted to firms with at least 100 patents in their portfolio. None of the top 10 firms with the highest level of *tech differentiation* ranks in the top 10 identified by the other measures.

Table A.6: Top 10 firms with the most differentiated technology portfolio for selected industries and years

Industry/year	Rank	Text	Classes	Subclasses	Citations
Automobiles and trucks (SIC37, 2012)	1	Gentex	Autoliv	Federal-Mogul Holdings	American Axle & MFG
	2	American Axle & MFG	American Axle & MFG	Autoliv	Borgwarner
	3	Autoliv	Borgwarner	American Axle & MFG	Tenneco
	4	Tenneco	Federal-Mogul Holdings	Borgwarner	Tesla
	5	Tesla	Tenneco	Gentex	Federal-Mogul Holdings
	6	Federal-Mogul Holdings	Tesla	Tenneco	Gentex
	7	Borgwarner	TRW Automotive Holdings	TRW Automotive Holdings	Autoliv
	8	TRW Automotive Holdings	Textron	Tesla	Rockwell Collins
	9	Ford Motor	Ford Motor	Textron	Visteon
	10	Raytheon Technologies	Raytheon Technologies	Goodrich	TRW Automotive Holdings
Chemicals and allied products (SIC28, 1983)	1	Pharmacia & Upjohn	INTL Flavors & Fragrances	INTL Flavors & Fragrances	INTL Flavors & Fragrances
	2	INTL Flavors & Fragrances	Smithkline Beckman	Smithkline Beckman	Petrolife
	3	Smithkline Beckman	Pharmacia & Upjohn	Pharmacia & Upjohn	Smithkline Beckman
	4	Squibb	Squibb	Petrolife	Cordant Technologies
	5	Schering-Plough	Schering-Plough	Cordant Technologies	Ex-Cell-O
	6	Wyeth	Wyeth	Ex-Cell-O	Pharmacia & Upjohn
	7	Lilly (Eli) & Co	Lilly (Eli) & Co	NL Industries	Schering-Plough
	8	Merck & Co	Pfizer	Schering-Plough	Uniroyal
	9	Pfizer	Merck & Co	Squibb	Hercules
	10	Petrolife	Warner-Lambert	Newmarket	Quantum Chemical
Electronic and other electrical equipment and components (SIC36, 2015)	1	Universal Display	Remy International	Amphenol	Remy International
	2	INTL Rectifier	Amphenol	Universal Display	Universal Display
	3	Intermolecular	Whirlpool	Remy International	Amphenol
	4	Alpha & Omega Semiconductor	Cirrus Logic	Whirlpool	Hubbell
	5	Amkor Technology	Acuity Brands	Hubbell	L3 Technologies
	6	First Solar	Power Integrations	Sunpower	Sunpower
	7	Remy International	Hubbell	First Solar	Adtran
	8	Sunpower	Synaptics	Intermolecular	Maxlinear
	9	Digitmarc	Sunpower	Omnivision Technologies	Intermolecular
	10	Maxlinear	Alpha & Omega Semiconductor	Power Integrations	Whirlpool
Steam, gas, and hydraulic turbines, and turbine generator set units (SIC35, 2005)	1	Metrologic Instruments	Kennametal	Timken	Sauer-Danfoss
	2	Immersion	Donaldson	Sauer-Danfoss	Kennametal
	3	Diebold Nixdorf	Brunswick	Cymer	Immersion
	4	Timken	Cymer	Kennametal	Timken
	5	Kennametal	Timken	Tecumseh Products	Cymer
	6	Sauer-Danfoss	Metrologic Instruments	Brunswick	Brunswick
	7	Electronics for Imaging	Sauer-Danfoss	Black & Decker	FMC Technologies
	8	Cymer	Lexmark INTL	Metrologic Instruments	Tecumseh Products
	9	Donaldson	Quantum	Donaldson	Nordson
	10	Brunswick	Western Digital	Cameron International	Dover

Notes: The table ranks the top 10 firms with the most unique and differentiated technology portfolio in their industry in a given year by *tech differentiation*, *tech differentiation (class)*, *tech differentiation (subclass)*, and *tech differentiation (citation)* (industries defined by 2-digit SIC). The selection is restricted to firms with at least 100 patents in their portfolio. The average overlap between the top 10 firms identified by *tech differentiation* and those identified by *tech differentiation (class)*, *tech differentiation (subclass)* and *tech differentiation (citation)* are 0.63, 0.63, and 0.58 respectively.

Table A.7: Additional fixed effects Tobin's Q

	Tobin's Q								
	Industry*year fixed effects			Technology*year fixed effects			Both combined		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Tech differentiation	1.951***	1.627***	2.494***	3.028***	2.604***	2.194***	2.555***	1.804***	1.764***
	(0.416)	(0.457)	(0.594)	(0.477)	(0.490)	(0.574)	(0.486)	(0.548)	(0.637)
R&D intensity	0.837***	0.826***	0.313***	0.885***	0.890***	0.324***	0.854***	0.843***	0.302***
	(0.057)	(0.061)	(0.064)	(0.058)	(0.058)	(0.061)	(0.058)	(0.062)	(0.065)
Citation-weighted patents	0.106***	0.103***	0.073***	0.100***	0.096***	0.072***	0.102***	0.098***	0.071***
	(0.007)	(0.008)	(0.010)	(0.008)	(0.008)	(0.010)	(0.007)	(0.008)	(0.010)
Firm FE	No	No	Yes	No	No	Yes	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Industry*year FE	No	Yes	Yes	No	No	No	No	Yes	Yes
Technology FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Technology*year FE	No	No	No	No	Yes	Yes	No	Yes	Yes
Number of observations	38,247	38,247	38,247	37,973	37,973	37,973	37,973	37,973	37,973
Number of firms	4,049	4,049	4,049	3,996	3,996	3,996	3,996	3,996	3,996
Within r2			0.244			0.196			0.272
Between r2			0.064			0.076			0.066
Overall r2	0.191	0.255	0.082	0.166	0.199	0.081	0.198	0.281	0.091
	Marginal effects (%)								
Tech differentiation	7.34	6.09	9.48	11.63	9.92	8.30	9.73	6.78	6.62
R&D intensity	16.68	16.43	5.94	17.66	17.77	6.14	17.01	16.77	5.70
Citation weighted patents	24.21	23.41	16.20	22.68	21.79	15.82	23.09	22.20	15.56

Notes: The table reports coefficient estimates from an OLS regression (columns (1), (2), (4), (5), (7) and (8)) or a linear firm-fixed effects regression (columns (3), (6) and (9)). The sample is an unbalanced panel with firm fiscal years ranging from 1989 to 2015. As a result of missing values of firm performance indicators in some years, the number of observations varies across columns. *Tobin's Q* and *citation-weighted patents* are log transformed. Additional control variables include *Total assets (log), Firm age (log), Leverage, Cash, Asset tangibility, tech specialization,* and *prod market competition (log).* We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. Control variables are lagged by one year. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on 3-digit SIC. Technology fixed effects are generated based on the NBER technology subcategories of patents contained in the portfolio of firm *i* in year *t*. The given dummy equals one if at least one patent in the portfolio is assigned to the given NBER technology subcategory. Definitions of variables are provided in Table A.1 in Appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable caused by a one-standard deviation increase of the corresponding explanatory variable. *, **, and *** denotes significance at the 10%, 5%, and 1% level.

Table A.8: Additional fixed effects ROA

	ROA								
	Industry*year fixed effects			Technology*year fixed effects			Both combined		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Tech differentiation	1.082***	1.088***	0.633***	-0.343**	-0.390**	0.634***	0.097	0.012	0.492**
	(0.146)	(0.161)	(0.182)	(0.150)	(0.157)	(0.157)	(0.156)	(0.176)	(0.191)
R&D intensity	-0.912***	-0.923***	-0.461***	-0.946***	-0.949***	-0.466***	-0.913***	-0.919***	-0.464***
	(0.033)	(0.034)	(0.040)	(0.032)	(0.032)	(0.038)	(0.033)	(0.034)	(0.040)
Citation-weighted patents	0.005**	0.005*	0.007*	0.013***	0.012***	0.010***	0.011***	0.010***	0.010**
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Firm FE	No	No	Yes	No	No	Yes	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Industry*year FE	No	Yes	Yes	No	No	No	No	Yes	Yes
Technology FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Technology*year FE	No	No	No	No	Yes	Yes	No	Yes	Yes
Number of observations	38,375	38,375	38,375	38,097	38,097	38,097	38,097	38,097	38,097
Number of firms	4,049	4,049	4,049	3,995	3,995	3,995	3,995	3,995	3,995
Within r2			0.132			0.096			0.151
Between r2			0.367			0.537			0.379
Overall r2	0.466	0.495	0.319	0.463	0.475	0.413	0.478	0.515	0.333
	Marginal effects (%)								
Tech differentiation	3.93	3.95	2.30	-1.25	-1.42	2.31	0.35	0.04	1.79
R&D intensity	-16.85	-17.05	-8.51	-17.43	-17.49	-8.58	-16.83	-16.93	-8.56
Citation weighted patents	1.08	1.01	1.43	2.57	2.42	2.12	2.34	2.10	1.97

Notes: The table reports coefficient estimates from an OLS regression (columns (1), (2), (4), (5), (7) and (8)) or a linear firm-fixed effects regression (columns (3), (6) and (9)). The sample is an unbalanced panel with firm fiscal years ranging from 1989 to 2015. As a result of missing values of firm performance indicators in some years, the number of observations varies across columns. *Tobin's Q* and *citation-weighted patents* are log transformed. Additional control variables include *Total assets (log), Firm age (log), Leverage, Cash, Asset tangibility, tech specialization,* and *prod market competition (log)*. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. Control variables are lagged by one year. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on 3-digit SIC. Technology fixed effects are generated based on the NBER technology subcategories of patents contained in the portfolio of firm *i* in year *t*. The given dummy equals one if at least one patent in the portfolio is assigned to the given NBER technology subcategory. Definitions of variables are provided in Table A.1 in Appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable caused by a one-standard deviation increase of the corresponding explanatory variable. *, **, and *** denotes significance at the 10%, 5%, and 1% level.

Table A.9: Technology differentiation and firm performance by size of patent portfolio

	Tobin's Q		ROA	
	(1)	(2)	(3)	(4)
	Firm patent count <=4	Firm patent count >=60	Firm patent count <=4	Firm patent count >=60
Tech differentiation	5.710**	3.459***	1.667*	0.848***
	(2.751)	(0.733)	(0.910)	(0.177)
R&D intensity	0.322**	0.439***	-0.376***	-0.447***
	(0.153)	(0.117)	(0.134)	(0.070)
Citation-weighted patents	0.072**	0.062***	0.017	0.009**
	(0.033)	(0.014)	(0.012)	(0.004)
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
Number of observations	4,543	16,537	4,557	16,582
Number of firms	968	1,013	968	1,013
Within r2	0.188	0.198	0.105	0.110
Between r2	0.080	0.065	0.008	0.294
Overall r2	0.075	0.082	0.001	0.211
	Marginal effects (in %)			
Tech differentiation	7.50	13.98	2.11	3.21
R&D intensity	6.24	6.38	-7.08	-6.30
Citation weighted patents	7.18	11.45	1.59	1.64

Notes: The table reports coefficient estimates from a linear firm-fixed effects regression. For each firm, we calculate *firm patent count* as the total number of granted patents filed between 1984 and 2015. We select two subsamples of firms, firms with at most 4 patents (i.e. the 25^{th} percentile of *firm patent count*) and firms with at least 60 patents (i.e. the 75^{th} percentile of *firm patent count*), and run split sample regressions. Additional control variables include *Total assets (log), Firm age (log), Leverage, Cash, Asset tangibility, tech specialization,* and *prod market competition (log)*. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. Control variables are lagged by one year. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on 3-digit SIC. Definitions of variables are provided in Table A.1 in Appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable caused by a one-standard deviation increase of the corresponding explanatory variable. *, **, and *** denotes significance at the 10%, 5%, and 1% level.

Table A.10: Technology differentiation and firm performance for specialized firms (a single product market industry) versus diversified firms

	Tobin's Q		ROA	
	(1)	(2)	(3)	(4)
	Specialized firms (one industry)	Diversified firms (multiple industries)	Specialized firms (one industry)	Diversified firms (multiple industries)
Tech differentiation	4.451***	3.086***	1.334***	0.668***
	(0.937)	(0.690)	(0.281)	(0.166)
R&D intensity	0.263***	0.469***	-0.454***	-0.517***
	(0.073)	(0.107)	(0.045)	(0.069)
Citation-weighted patents	0.098***	0.063***	0.016***	0.003
	(0.014)	(0.012)	(0.006)	(0.004)
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
Number of observations	17,945	20,302	18,020	20,355
Number of firms	2,415	1,634	2,416	1,633
Within r2	0.163	0.171	0.074	0.093
Between r2	0.031	0.058	0.012	0.292
Overall r2	0.038	0.057	0.002	0.228
	Marginal effects (in %)			
Tech differentiation	14.10	12.90	3.96	2.62
R&D intensity	6.00	5.93	-10.10	-6.36
Citation weighted patents	19.75	14.71	2.92	0.70

Notes: The table reports coefficient estimates from a linear firm-fixed effects regression. The Securities and Exchange Commission's (SEC) requires firms to report business segments which exceed 10% of sales. We define specialized firms as firms which report only one and the same SIC3 industry during the entire observation window and diversified firms as firms which reported at least two SIC3 industries in a given year or switched between SIC3 industries over time. Additional control variables include *Total assets (log), Firm age (log), Leverage, Cash, Asset tangibility, tech specialization,* and *prod market competition (log)*. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. Control variables are lagged by one year. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on 3-digit SIC. Definitions of variables are provided in Table A.1 in Appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable caused by a one-standard deviation increase of the corresponding explanatory variable. *, **, and *** denotes significance at the 10%, 5%, and 1% level.

Table A.11: Alternative technology differentiation measures

	Tobin's Q		ROA	
	(1)	(2)	(3)	(4)
Tech differentiation (within industry)	0.330***		0.079***	
	(0.102)		(0.028)	
Tech differentiation (compared to closest firm)		0.439***		0.105***
		(0.086)		(0.029)
R&D intensity	0.317***	0.326***	-0.462***	-0.460***
	(0.059)	(0.059)	(0.037)	(0.037)
Citation-weighted patents	0.056***	0.076***	0.003	0.007**
	(0.008)	(0.009)	(0.003)	(0.003)
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
Number of observations	38,247	38,247	38,375	38,375
Number of firms	4,049	4,049	4,049	4,049
Within r2	0.157	0.158	0.075	0.076
Between r2	0.074	0.074	0.409	0.405
Overall r2	0.065	0.065	0.324	0.319
	Marginal effects (in %)			
Tech differentiation (within industry)	3.25		0.77	
Tech differentiation (compared to closest firm)		8.62		1.97
R&D intensity	6.02	6.19	-8.53	-8.49
Citation weighted patents	12.26	16.73	0.55	1.49

Notes: The table reports coefficient estimates from a linear firm-fixed effects regression. The sample is an unbalanced panel with firm fiscal years ranging from 1989 to 2015. As a result of missing values of firm performance indicators in some years, the number of observations varies across columns. *Tobin's Q* and *citation-weighted patents* are log transformed. *Tech differentiation (within industry)* of firm *i* in year *t* is calculated as one minus the average *tech similarity* between firm *i* and all other firms from the same SIC3 industry as firm *i* in year *t*, and *tech differentiation (compared to closest firm)* of firm *i* in year *t* is calculated as one minus the maximum *tech similarity* between firm *i* and all other firms in year *t*. Additional control variables include *Total assets (log), Firm age (log), Leverage, Cash, Asset tangibility, tech specialization,* and *prod market competition (log)*. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. Control variables are lagged by one year. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on 3-digit SIC. Definitions of variables are provided in Table A.1 in Appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable caused by a one-standard deviation increase of the corresponding explanatory variable. *, **, and *** denotes significance at the 10%, 5%, and 1% level.

Table A.12: Long-term effect of technology differentiation on Tobin's Q

	t	t+1	t+2	t+3	t+4	t+5
	(1)	(2)	(3)	(4)	(5)	(6)
Tech differentiation	3.469***	2.934***	2.611***	1.936***	1.380**	1.302**
	(0.552)	(0.577)	(0.583)	(0.588)	(0.605)	(0.627)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	38,247	33,608	29,620	26,115	23,109	20,458
Number of firms	4,049	3,845	3,500	3,128	2,816	2,456
Within r2	0.159	0.160	0.135	0.125	0.123	0.124
Between r2	0.076	0.065	0.045	0.029	0.020	0.011
Overall r2	0.068	0.059	0.047	0.036	0.026	0.017
Marginal effects of tech differentiation (in %)	13.42	11.36	10.15	7.50	5.34	5.07

Notes: The table summarizes the long-term relationship between *tech differentiation* of firm *i* in year *t* and *Tobin's Q* measured in year *t* up to *t+5* by linear firm-fixed effects regression. The sample is an unbalanced panel with firm fiscal years ranging from 1989 to 2015. As a result of missing values, the number of observations vary across columns. *Tobin's Q* is log transformed. Control variables include *R&D intensity*, *Citation-weighted patents (log)*, *Tech specialization*, *Total assets (log)*, *Firm age (log)*, *Leverage*, *Cash*, *Asset tangibility*, and *Prod market competition (log)*. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on 3-digit SIC. Definitions of variables are provided in Table A.1 in Appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable caused by a one-standard deviation increase of *tech differentiation*. *, **, and *** denotes significance at the 10%, 5%, and 1% level.

Table A.13: Long-term effect of technology differentiation on ROA

	t	t+1	t+2	t+3	t+4	t+5
	(1)	(2)	(3)	(4)	(5)	(6)
Tech differentiation	0.908***	0.872***	0.805***	0.784***	0.572***	0.206
	(0.149)	(0.162)	(0.172)	(0.174)	(0.181)	(0.187)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	38,375	33,723	29,720	26,210	23,197	20,537
Number of firms	4,049	3,849	3,501	3,128	2,819	2,460
Within r2	0.076	0.033	0.026	0.026	0.027	0.027
Between r2	0.407	0.142	0.056	0.002	0.005	0.013
Overall r2	0.318	0.130	0.056	0.002	0.004	0.014
Marginal effects of tech differentiation (in %)	3.30	3.20	2.98	2.93	2.15	0.78

Notes: The table summarizes the long-term relationship between *tech differentiation* of firm *i* in year *t* and *ROA* measured in year *t* up to *t+5* by linear firm-fixed effects regression. The sample is an unbalanced panel with firm fiscal years ranging from 1989 to 2015. As a result of missing values, the number of observations vary across columns. Control variables include *R&D intensity*, *Citation-weighted patents (log)*, *Tech specialization*, *Total assets (log)*, *Firm age (log)*, *Leverage*, *Cash*, *Asset tangibility*, and *Prod market competition (log)*. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on 3-digit SIC. Definitions of variables are provided in Table A.1 in Appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable caused by a one-standard deviation increase of *tech differentiation*. *, **, and *** denotes significance at the 10%, 5%, and 1% level.

Table A.14: Technology differentiation and firm performance by industry type

	Industries with high versus low R&D intensity				Industries with high versus low product market rivalry (10-K business description based)				Industries with high versus low product market rivalry (Compustat-based HHI)				Industries with high versus low product market rivalry (Census-based HHI)			
	Tobin's Q		ROA		Tobin's Q		ROA		Tobin's Q		ROA		Tobin's Q		ROA	
	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Tech differentiation	3.934***	1.658**	0.957***	0.554***	4.602***	1.132*	1.291***	0.293*	3.770***	-1.376	0.965***	0.054	3.177***	1.942	0.868***	0.708*
	(0.686)	(0.683)	(0.188)	(0.168)	(0.776)	(0.643)	(0.214)	(0.152)	(0.577)	(1.349)	(0.156)	(0.336)	(0.624)	(1.217)	(0.166)	(0.422)
R&D intensity	0.295***	0.876***	-0.456***	-0.507**	0.312***	0.464**	-0.449***	-0.444***	0.322***	0.574	-0.454***	-0.682**	0.323***	0.995***	-0.469***	-0.672**
	(0.062)	(0.273)	(0.038)	(0.240)	(0.065)	(0.180)	(0.040)	(0.141)	(0.060)	(0.361)	(0.038)	(0.304)	(0.062)	(0.361)	(0.039)	(0.334)
Citation-weighted patents	0.090***	0.019	0.009**	0.003	0.095***	0.038***	0.013***	-0.000	0.080***	0.052*	0.010***	0.003	0.079***	0.036	0.006*	0.007
	(0.011)	(0.012)	(0.004)	(0.003)	(0.012)	(0.011)	(0.004)	(0.004)	(0.010)	(0.027)	(0.003)	(0.009)	(0.010)	(0.023)	(0.004)	(0.006)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	29,702	8,545	29,806	8,569	24,912	13,331	25,013	13,358	35,541	2,706	35,674	2,701	31,453	3,183	31,544	3,188
Number of firms	3,357	975	3,358	973	2,920	1,442	2,920	1,442	3,854	347	3,855	344	3,379	379	3,379	378
Within r2	0.165	0.196	0.077	0.074	0.173	0.164	0.075	0.069	0.160	0.183	0.074	0.087	0.146	0.173	0.077	0.105
Between r2	0.081	0.018	0.437	0.035	0.068	0.046	0.469	0.117	0.073	0.006	0.421	0.289	0.089	0.097	0.516	0.037
Overall r2	0.073	0.010	0.347	0.012	0.070	0.028	0.377	0.078	0.067	0.010	0.329	0.152	0.074	0.044	0.394	0.008
	Marginal effects (in %)															
Tech differentiation	15.53	5.95	3.51	1.93	18.00	4.24	4.64	1.08	14.67	-4.86	3.50	0.20	11.92	8.78	3.07	3.07
R&D intensity	5.99	6.40	-9.03	-3.59	6.62	4.62	-9.24	-4.32	6.24	5.35	-8.57	-6.08	6.45	9.61	-9.08	-6.29
Citation weighted patents	20.17	3.94	1.94	0.62	21.72	7.84	2.77	-0.04	17.73	10.47	2.05	0.66	17.03	8.80	1.29	1.71

Notes: The table reports coefficient estimates from a linear firm-fixed effects regression. The sample is an unbalanced panel with firm fiscal years ranging from 1989 to 2015. As a result of missing values of firm's performance indicators in some years, the number of observations varies across columns. The sample is split by the mean of *industry R&D intensity* (columns 1-4), by the mean of *industry prod market competition* (columns 5-8), by the mean of *industry prod market competition (Compustat-based HHI)* (column 9-12), and by the mean of *industry prod market competition (U.S. Census-based HHI)* (column 13-16). *Tobin's Q* and *citation-weighted patents* are log transformed. Additional control variables include *Total assets (log)*, *Firm age (log)*, *Leverage*, *Cash*, *Asset tangibility*, *Tech specialization*, and *Prod market competition (log)*. Control variables are lagged by one year. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on 3-digit SIC (column 1-12) or 3-digit NAICS (column 13-16). Definitions of variables are provided in Table A.1 in Appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable caused by a one-standard deviation increase of the corresponding explanatory variable. *, **, and *** denotes significance at the 10%, 5%, and 1% level