Research Policy 51 (2022) 104426

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

A language-based approach to measuring creative exploration

Vladimir A. Gatchev *, Christo A. Pirinsky, Buvaneshwaran Venugopal

College of Business Administration at the University of Central Florida, 12744 Pegasus Dr., Orlando, FL 32816-1400, United States

ARTICLE INFO

JEL classification:
D83
G12
G32
O30

Keywords:
Exploration
Learning by doing
Innovation
Textual analysis, Firm risk
Firm value

ABSTRACT

We propose a new measure of the exploratory activities of companies based on the idea that experimentation with new courses of action and the need to describe them entails the adoption of new words in firm regulatory disclosures. Unlike traditional indicators, such as R&D spending, the proposed exploration indicator is available for all publicly traded firms across all industries. The exploration indicator predicts firm knowledge accumulation, as measured by future patenting and trademarking activities. It further shows that firm exploration declines after periods of high R&D spending and over time. The exploration indicator correlates positively with firm risk and exhibits a distinct positive impact on firm value unexplained by traditional innovation indicators. Our language-based approach can be applied to measure creative contributions in other domains, such as government grant applications and academic publications.

Every aspect of the life of a people is reflected in the words they use to talk about themselves and the world around them. As their world changes – through invention, discovery, revolution, evolution, or personal transformation – so does their language.

John Algeo, Fifty Years among the New Words

1. Introduction

Exploration, or the tendency of individuals and organizations to experiment with new courses of action, has received a great deal of attention in numerous disciplines, ranging from economics to organizational science and management.[1] More specifically, exploration has been recognized as an important factor for the accumulation of knowledge, early-stage innovation, and economic development (Arrow, 1962; Manso, 2011; Kerr et al., 2014; Lucas Jr and Moll, 2014).[2] Yet, we still know relatively little about this early stage of the creative process, possibly because exploration is difficult to measure (Dziallas and Blind, 2019).

Traditional indicators of exploration, such as Research and Development (R&D) spending, while highly useful, also exhibit some shortcomings.[3] For example, in many important sectors of the economy (e.g., Energy, Finance, Transportation), most firms do not record R&D spending. Moreover, while R&D spending usually accounts for some activities related to the introduction of new products and production processes, exploration spans a wider and more complex range of activities also related to the creation and implementation of new production processes, the exploration of new markets, the adoption of new supply chain structures, and the introduction of new organizational structures (Schumpeter, 1934).[4]

In this paper, we propose a general methodology for measuring creative exploration. Our methodology is applicable to a wide range of firms and a wide range of settings and is based on the idea that experimentation with new concepts and practices is reflected in natural language. As firms experiment with new courses of action, they need to use new vocabulary to describe their activities. Innovations in business

* Corresponding author.
E-mail address: vgatchev@ucf.edu (V.A. Gatchev).

[1] See, for example, Arrow (1962), March (1991), Irwin and Klenow (1994), Foster and Rosenzweig (1995), Jovanovic and Nyarko (1996), Benner and Tushman (2003).

[2] Throughout the paper, we use the terms "exploration" and "experimentation" interchangeably.

[3] A large literature has emphasized that spending on R&D enhances innovation output and the accumulation of technological knowledge (Romer, 1986; Grossman and Helpman, 1991; Aghion and Howitt, 1992; 1998a; Akcigit and Kerr, 2018).

[4] Many researchers have also acknowledged that the definition of industrial R&D may be too restrictive (see among others Freeman and Soete (2009) and the references therein) and that much can be learned by using a range of measures related to the innovative activities of firms (Pavitt, 1982).

https://doi.org/10.1016/j.respol.2021.104426
Received 10 October 2020; Received in revised form 28 October 2021; Accepted 2 November 2021
Available online 12 November 2021

vocabulary could be achieved either by the adoption of new meanings for existing words or by the creation of brand-new words. Consistent with the latter is the fact that advancements in science and technology tend to be among the most prolific sources of neologisms (Dyke, 1992; Algeo, 1993; Crystal, 2002; McDonald, 2005).

We apply our methodology to mandated firm disclosures and construct an annual indicator of exploratory activity for all publicly traded firms in the U.S. between 1997 and 2017. Our study is made possible by existing regulations in the U.S., which mandate all publicly traded firms to timely and accurately disclose in their Form 10-K annual filings to the Securities and Exchange Commission (SEC) any material information related to their business.[5] The annual filings start with a business description section containing detailed information about the firm's products and business practices. We start by extracting the business description sections from all filings and summarizing their content based on all nouns and proper nouns. We then identify the subset of all nouns and proper nouns that are used by a firm for the first time in a given year, relative to the firm's history of annual filings. To exclude routine words, we focus only on the subset of new words with relatively rare usage in the past by the firm's peers and refer to them as "explorative words." Our exploration indicator at the firm-year level is constructed by aggregating all explorative words for each firm in each year.

The central idea of our study is that the new words each firm uses when describing its business in the annual filings denote possible exploratory activities during the year. Thus, we use these indications of exploratory activities to construct our exploration indicator. For a new exploration indicator to be valid, it needs to relate positively to real exploratory activities and the accumulation of knowledge. In addition, for a new indicator to be able to complement and extend already existing indicators, the information it provides about the exploratory activities of firms should not be fully captured by other indicators.

We perform validation tests and show that our proposed exploration indicator is positively related to other proxies of exploratory activities by firms, such as R&D expenditures and involvement in new technologies. Our exploration indicator exhibits a positive and significant relation with R&D spending, both across and within industrial sectors, and predicts future R&D spending at the level of individual firms. As another validation test, we examine how the exploration indicator relates to firm involvement in technologies from the list of emerging new technologies developed by Gartner Inc. Gartner is a leading IT consulting firm that conducts annual comprehensive surveys to create a qualitative ranking of notable new technologies expected to become mainstream. We show that more explorative firms, as defined by our indicator, are more likely to work on cutting-edge technologies when compared with less explorative firms.[6]

Economists have recognized that the accumulation of knowledge is fundamentally linked to experimentation (Romer, 1994; Foster and Rosenzweig, 1995; Aghion and Howitt, 1998b). On the one hand, the emergence of new technologies and business practices needs to follow a steadily evolving process of experimentation, given that "learning associated with repetition of essentially the same problem is subject to sharply diminishing returns" (Arrow, 1962). On the other hand, the accumulation of knowledge could also stimulate future exploratory activities by providing firms with the know-how and a wider set of opportunities.

To see how our exploration indicator relates to knowledge accumulation, we examine the dynamic relations of our indicator with the number of citation-weighted patent applications and grants and the number of trademark applications and grants at the firm level.[7] Patents are commonly used as an indicator of the stock of knowledge and innovative output by firms (Griliches, 1990; Bloom and Van Reenen, 2002; Hall et al., 2005). Some recent studies have also focused on trademarks as an alternative indicator of innovation output (Mendonça et al., 2004; Dinlersoz et al., 2019; Faurel et al., 2021). We find that our exploration indicator predicts both patent and trademark *applications* as well as patent and trademark *grants* over the next three years. Consistent with the idea that our indicator captures early-stage exploration efforts, we also find that patent and trademark *grants* do not predict an increase in future explorations.

To validate our exploration indicator in a time-series context, we examine how firm propensity for exploration changes with firm age. We expect exploration to decline with firm age for at least two reasons. First, every firm is expected to strengthen its position in the marketplace over time. As a consequence, older and established organizations would find it less beneficial to experiment with new courses of action (Hobijn and Jovanovic, 2001). Second, older organizations could find it more difficult to make radical changes in their operation and strategy due to organizational inertia (Hannan and Freeman, 1984; Nelson and Winter, 1982). Consistent with this prediction, we find that our exploration indicator declines with firm age.

March (1991) points out that, compared with the returns from exploitation (i.e., on-going operations), the returns from exploration are more remote in time and less certain. Thus, an exploration indicator should be able to capture the uncertainties inherent in all exploratory activities and should correlate positively with firm risk. Because firm uncertainty is expected to be reflected in equity prices, we measure risk based on firm stock returns (Carpenter and Petersen, 2002; Pastor and Veronesi, 2009). We show that more explorative firms tend to be riskier than less explorative firms in terms of both systematic and idiosyncratic risk.

For our final analysis, we demonstrate the usefulness of our indicator by examining an important question: Do exploratory activities add economic value? To answer this question, we employ the approach in Hall et al. (2005) and construct an indicator of cumulative stock of exploration. We then examine how firm value, as measured by the ratio of the firm's market value of assets to book value of assets (often denoted as *Q*), relates to the exploration stock indicator, while making sure to control for traditional variables, such as the stock of R&D and patents. We find a positive and significant relation between our exploration stock variable and firm value, a relation that is not subsumed by R&D and patenting activities. The relation between our exploration stock variable and firm value is economically meaningful – a one standard deviation increase in the exploration stock variable is associated with an increase in firm value by around five percent, on average.

We present extensive evidence that our proposed exploration indicator is related to firm early exploratory activities.[8] Moreover, the indicator captures relevant information above and beyond the information captured by R&D spending. While in many important sectors of the economy most firms do not report any R&D spending, our variable can serve as an indicator of potential exploratory efforts in all areas of economic activity. Indeed, our exploration indicator is positive for 89

[5] Firms must file Form 10-K annually with the SEC pursuant to Sections 13 or 15(d) of the Securities Exchange Act of 1934, as amended. These reports give a comprehensive picture of the business and financial conditions of firms and include audited financial statements. Throughout the paper we refer to Form 10-K filings also as "annual filings."

[6] These tests are introduced in greater detail in Section 3.

[7] As we discuss in Section 2.1, there is a time gap of around 3.1 (1.8) years between application dates and grant dates for patents (trademarks). Section 3.3.1 describes the calculation of the citation-weighted patent variable.

[8] Our validation results also suggest that it is unlikely that our language-based measure reflects entirely empty rhetoric or symbolisms (i.e., spin-writing). Such possibility could be largely mitigated by (i) the fact that firms and top managers face reputational and legal costs for purposefully misrepresenting their activities in Form 10-K annual filings and (ii) the construction of our measure, which includes only new words that have not been widely used by other firms in the industry and thus are less likely to be attention-grabbing and to be used strategically by firms.

percent of all publicly traded firms in a given year and is available across all industries. In contrast, more than 60 percent of all publicly traded firms do not report R&D spending in a given year, with some industries (e.g., Finance) reporting trivial R&D spending. Furthermore, while R&D spending is predominantly linked to the development of new products and production processes, our variable captures a broader set of exploratory activities. Consistent with this idea, we confirm that our exploration indicator spans various activities ranging from the development of new products and the introduction of new production processes to the composition of the firm supply chain and its organizational structure.

Traditionally, economists have relied predominantly on hard information for quantitative analysis. However, advancements in computing have allowed researchers to extract information from unstructured data such as text, images, and videos (e.g., Henderson et al., 2012; Blankespoor et al., 2017). An increasing number of studies use textual analysis to capture economic phenomena that are difficult to measure using hard information. For example, Hoberg and Phillips (2016) use firm regulatory filings to define a new classification of industries by evaluating product similarities across firms. Kelly et al. (2021) use textual analysis of patent documents to quantify the level of patent novelty. Hassan et al. (2019) use transcripts of quarterly earnings calls to assess perceived political risk and its impact on firm performance. In this paper, we use textual analysis to identify words that could be associated with the early-stage exploratory activities of firms. We note, however, that our approach is general and could be applied to measure creative contributions in other domains, such as academic publications and government grant applications.

2. A language-based approach to measuring exploration

2.1. Data sources

We use multiple data sources, including Form 10-K annual filings by firms to the Securities and Exchange Commission (SEC), the Center for Research in Security Prices (CRSP), Compustat, the patent data of Kogan et al. (2017), and federal trademark applications from the United States Patent and Trademark Office (USPTO).

We compile our main sample using all publicly traded firms in CRSP and Compustat with fiscal years between 1996 and 2017. CRSP provides historic information at daily and monthly frequencies on prices, returns, and shares outstanding (among other data) for all stocks traded on the U. S. major exchanges: NYSE, AMEX, and Nasdaq. Compustat is supplied by Standard & Poor's and provides comprehensive accounting data at quarterly and annual frequencies for all publicly traded U.S. companies. For this study, we use annual Compustat data that directly corresponds to the Form 10-K filings of firms.

Years are defined based on the date of the fiscal year-end so that fiscal years ending from January to May of year t are classified as year t–1 while fiscal years ending from June to December of year t are classified as year t. We require firms to have positive sales and assets of more than $1 million and keep only firms incorporated and located in the United States of America. The resulting sample contains 97,390 firm-year observations.

We identify Form 10-K annual filings for 94,334 firm-years between 1996 and 2017, which constitutes 96.9% of our initial sample. The coverage between 1997 and 2017 varies from 93.3% for the fiscal year 2002 to 99.3% for the fiscal year 2014. We also use annual filings for the fiscal year 1996, whenever available. The coverage for 1996 is relatively lower, at 79.3%. To identify explorative words as of year t, we require our sample firms to have at least one past year of annual filings. This last restriction results in a sample of 88,175 firm-year observations between 1997 and 2017 with a total of 11,195 unique firms. On average, the sample contains 4,199 firms per year with a maximum of 6,044 firms in 1998 and a minimum of 3,113 in 2017.

To examine the relation between our language-based exploration indicator and patent-based innovation indicators, we use the patent data of Kogan et al. (2017).[9] The data was constructed using United States Patent and Trademark Office (USPTO) patent text files obtained from Google. The data contains information on 2,950,305 patent applications filed between 1926 and 2019 by 8,509 companies covered in CRSP. We distinguish between patent application dates and patent grant dates and construct two distinct patenting activity indicators: one measuring the number of patent applications in a given year and the other measuring the number of patent grants in a given year.

We also obtain data on federal trademark applications and granted registrations by listed companies from USPTO's website.[10] This dataset contains information on around 8.5 million federal trademark applications and granted registrations in the U.S. between 1870 and 2018. For each application, we can observe the filing date and the name of the entity applying for the trademark. Using a tri-gram vector decomposition method, we identify 21,384 listed companies in the trademark dataset which are also covered in Compustat. These companies filed 930,807 trademark applications and were granted 621,732 trademark registrations. We again distinguish between application dates and grant dates and construct two distinct trademark activity indicators: one measuring the number of trademark applications in a given year and the other measuring the number of trademark grants in a given year.

Based on the above information, we create a panel dataset containing the number of patent and trademark applications and grants by each firm, each year (see Section 3.3 for more details). On average, it takes around 3.1 years for a patent application to be approved and granted. The corresponding time for trademarks is around 1.8 years. The significant time gap between application and grant dates for both patents and trademarks allows us to examine the dynamic relations between our exploration variable and the two innovation indicators. In these analyses, we make sure also to consider the importance of R&D spending.

We obtain additional firm-level data needed for our analysis from the Compustat annual files and the CRSP daily files. All Compustat variables are measured at the end of each fiscal year. Compustat also provides Global Industrial Codes (GICs), based on an industry classification system developed by Morgan Stanley Capital International (MSCI) and Standard & Poor's. We combine some of the 24 GICS industry groups into 17 industry groups as follows: (i) Automobiles & Components (GIC: 2510) and Consumer Durables & Apparel (GIC: 2520) into Consumer Goods; (ii) Food & Staples Retailing (GIC: 3010), Food, Beverage & Tobacco (GIC: 3020), and Household & Personal Products (GIC: 3030) into Consumer Staples; (iii) Banks (GIC: 4010), Diversified Financials (GIC: 4020), and Insurance Companies (GIC:4030) into Financials; (iv) Technology Hardware & Equipment (GIC: 4520) and Semiconductors & Semiconductor Equipment (GIC: 4530) into IT Hardware & Semiconductors; and (v) Telecommunication Services (GIC: 5010) and Media & Entertainment (GIC: 5020) into Communications & Entertainment.

We express all dollar amounts in 2018 U.S. dollars using the GDP price deflator from the Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/series/GDPDEF). We use data from the CRSP daily files to compute the weekly returns of each stock in our sample. Moreover, CRSP directly provides the daily returns on an equally-weighted portfolio of all stocks in the database. We use these daily returns to construct weekly returns of the CRSP equally-weighted portfolio.

2.2. Summarizing the business descriptions in firm annual filings

To construct our language-based exploration indicator, we extract the business descriptions (item 1 or item 1A) from all Form 10-K annual filings in our sample. The business description section of the annual

[9] The data can be downloaded from https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth-Extended-Data.

[10] See link: https://www.uspto.gov/trademarks-application-process/checking-application-status-view-documents/trademark-bulk-data.

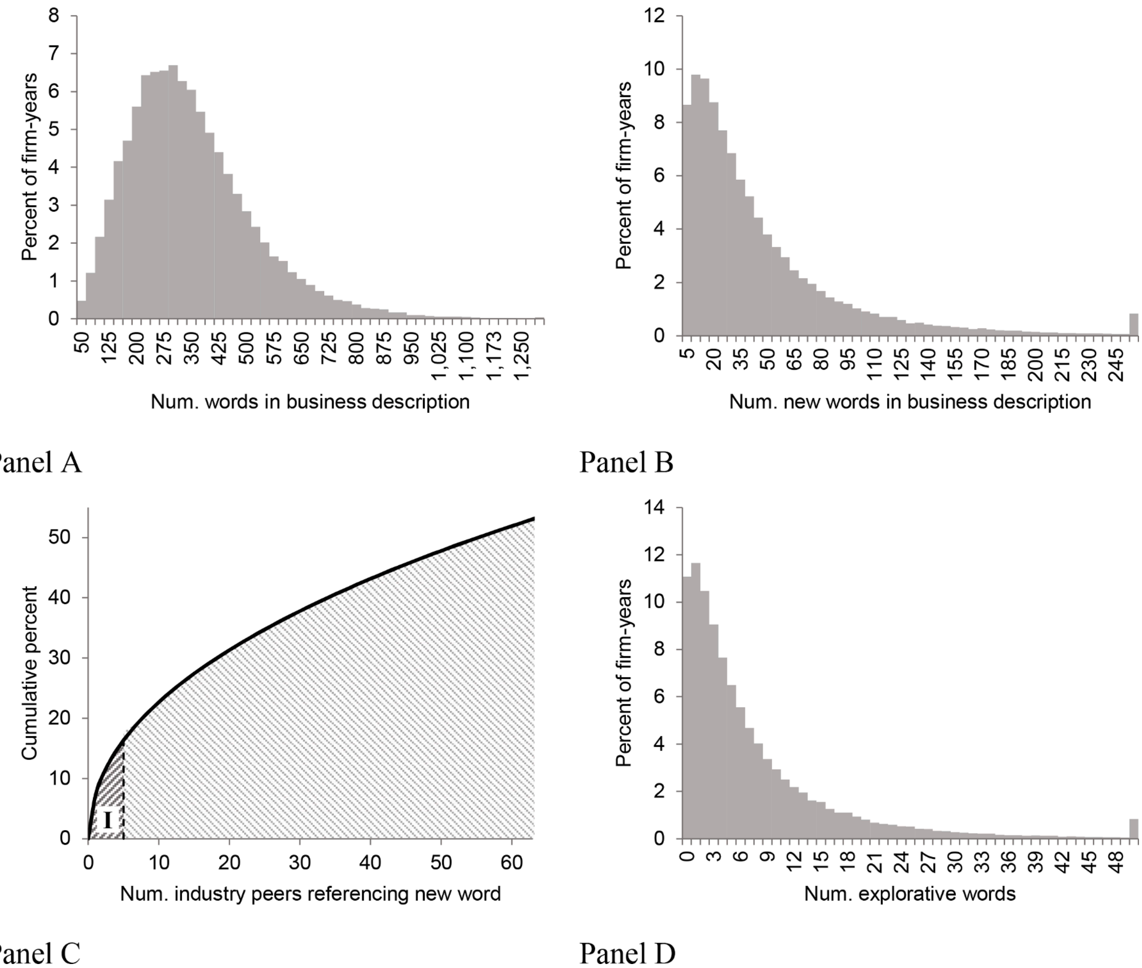

Fig. 1. The distribution of nouns and proper nouns and the distribution of explorative words within firm annual filings.

filing is particularly suited for our objectives because it contains detailed information on firm products, production processes, and major markets and supply sources. These disclosures are mandated by Item 101 of Regulation S-K and need to be accurate and up-to-date, which allows us to measure changes over time.[11]

Our first step is to summarize the business practice of each firm in terms of all noun and proper noun words used in its business description. To accomplish this, we start by dividing each business description into sentences.[12] Within each sentence, we identify noun and proper noun words based on a three-step approach. In the first step, we use Python's Natural Language Toolkit (NLTK) to classify each word into a grammatical part of speech (noun, pronoun, verb, etc.). In this step, we keep only words that either (i) are classified as nouns or proper nouns based on the word's context within the sentence (using Python's NLTK POS tagger) or (ii) can only be nouns, according to the Princeton University's WorldNet® lexical database.[13] In the second step, we bifurcate the resulting sample based on whether a word can be found in a major dictionary. In the third step, we manually review all words identified in the first step as nouns that are not in the dictionary and all words identified as proper nouns to filter out words with typos, personal names, and common abbreviations (e.g., USD, FY2010, and ROI).[14] We keep two-letter words (e.g., 5 G) and hyphenated words (e.g., E-tegrity). Following this three-step approach, we build a dictionary of all noun and proper noun words that were used in the business description sections of the annual filings of all publicly traded firms over the past 21 years.

We exclude words referenced in more than 25 percent of all business descriptions because such words are too generic and are unlikely to reflect exploration. Further, to be classified as an explorative word, we require a word to be used in at least two filings in our entire dataset, either by the same firm or by different firms. Finally, we exclude filings with fewer than 30 noun or proper noun words because such filings are unlikely to be informative (Hoberg and Phillips, 2016.

Our approach is general and captures a diverse set of firm activities. In addition to products (e.g., smartphone), our business characterization also includes brand names (e.g., Fix-a-Flat), production processes (e.g.,

[11] For instance, firms are required to describe in their Form 10-K annual filings "(t)he principal products produced and services rendered," "methods of distribution" and "(t)he sources and availability of raw materials" (https://www.ecfr.gov/cgi-bin/text-idx?amp;node=17:3.0.1.1.11&rgn=div5#se17.3.229_1101).

[12] The beginning of each paragraph is treated as a beginning of a new sentence. Within paragraphs, the end of one sentence and the beginning of the next sentence is identified by a combination of a period, a space, and a word beginning with a large cap letter. Cases involving "Mr.", "Mrs.", or "Ms." are excluded when identified as standalone sentences.

[13] WorldNet® is a large lexical database of words in the English language that groups words based on both their similarity of meaning and their semantic similarity (Princeton University, 2010).

[14] We also exclude generic words that appear in multiple contexts within different annual filings (e.g., "gse"). To facilitate this process, we use extensively *directEdgar* to manually identify each word within the actual filings.

fracking, lightweighting), business strategies (e.g., gamification), geographic locations, acquisitions, and ventures with other firms. To illustrate the context in which these words appear, let's take for example "gamification." It appears originally in the 2012 10-K filing of *One World Holdings, Inc* as follows: "Gamification: In order to stay competitive in the emerging realm of computer and online games, we plan to develop games that will incorporate the Prettie Girls characters into alternative counter-reality worlds, allowing girls to become their favorite doll and create online environments where they can live out their fantasies."

Fig. 1, Panel A, plots the frequency distribution of the number of all noun and proper noun words in the business descriptions of firm annual filings. The typical filing contains around 300 noun or proper noun words with some filings containing as many as 1000 words or more. These numbers are higher than the numbers reported in Hoberg and Phillips (2016), where the typical filing contains roughly 200 unique nouns and proper nouns. The reason for the higher number of words in our study is because we use a context-based approach to identify nouns that may or may not be in a major dictionary, whereas Hoberg and Phillips (2016) consider only nouns as identified by a major dictionary. Given that our study focuses on exploration and many new words do not enter formal dictionaries immediately, we believe the less restrictive approach we employ here is necessary.

2.3. Defining the exploration indicator

The point of departure in this study is that exploratory activity is reflected in natural language. The previous sub-section described our approach of identifying the list of all noun and proper noun words in the business description section of the annual filing of firm i in year t. We denote this list of words with $P_{i,t}$. Next, we identify the subset of new words ($\Delta P_{i,t}$) in a given year t as the words that were used by the firm for the first time in that year:

$$\Delta P_{i,t} = \left\{ \omega \middle| \omega \in P_{i,t}, \omega \notin \bigcup_{k=1}^{\infty} P_{i,t-k} \right\}. \quad (1)$$

The distribution of the number of new words per filing is plotted in Fig. 1, Panel B. We observe that the typical firm references around 29 new noun and proper noun words in a given year. The distribution is highly skewed with some firms introducing as many as 150–200 new words.

Our general approach of measuring firm-level exploration in a given year is to count the number of new words that could represent *novel* business activities each firm introduces for the first time in its history. To measure the novelty of an activity, we count the number of other firms in the same industry group that have already used the word characterizing this activity up to and including the current year. Specifically, for each new word (ω) from the set of new words ($\Delta P_{i,t}$) of a firm in a given year, we calculate the number of firms ($NF_{i,t}(\omega)$) from the same industry group that has used the same word in their business descriptions up to and including the current year. We define peers at the level of industry groups because words could have different meanings across different fields.[15]

All new words used by a firm for the first time in a given year could be classified into two groups depending on how many times these words have been referenced in the market. The first set of words includes all new words introduced by a firm that are also largely new to the market. We refer to this set of words as *explorative words*. The second set contains words that have been widely used in the market and thus are unlikely to represent explorations. Our analysis excludes the second set of words and is based on the set of *explorative words*.

To determine formally the two groups of new words, we build on Rogers (1962) who applies an *S*-curve adoption framework to examine how new ideas and products spread through a system of adopters.[16] Based on a normal adopter distribution, Rogers (1962) proposes using the 2.5th, 16th, 50th, and 84th percentiles to categorize the different stages of adoption. These percentiles are constructed using one and two standard deviations around the mean of a normal distribution. We follow this common approach and classify each new word of firm i in year t as an explorative or non-explorative word using the 16th percentile of $NF_{i,t}(\omega)$ (i.e., the number of firms from the same industry group that have used the same word in their business description up to and including the current year).[17]

Fig. 1, Panel C, plots the cumulative distribution of all new words introduced by a firm conditional on the number of industry peers referencing the word up to and including the current year. Around 16 percent of all new words are referenced by five or fewer industry peers (area I). As noted above, we focus our analysis on these relatively less common words, which we consider to be explorative words. Some examples of explorative words that were subsequently adopted in the IT Software & Services sector and the Materials sector are provided in Table A.2 of the Appendix. As discussed above, the identified explorative words capture a wide range of firm activities. In addition to products and processes, some explorative words are related to new technologies (e.g., blockchain), some to marketing aspects (e.g., millennial), and some to organizational attributes (e.g., onboarding).

Our firm-level exploration indicator for firm i in year t is constructed by counting the total number of explorative words introduced by firm i in year t. Fig. 1, Panel D, plots the frequency distribution of our proposed exploration indicator. The typical firm in our sample introduces around five explorative words each year. Similar to the total word count and the new word count variables, the distribution of our exploration indicator is also positively skewed, with only a small percentage of the firms introducing 25 or more explorative words in a given year.

2.4. Frontiers of the exploration indicator

Corporations engage in many types of innovation activities related to R&D, marketing and branding, software development, acquisition of tangible assets, and other aspects of a firm's business (see, for example, OECD, 2018, §4.2). The proposed exploration indicator captures a broad cross-section of these different types of innovation activities, including some activities related to R&D.

To better understand the frontiers of our exploration indicator, we need a full critical assessment of its relation to R&D activities, which are commonly used to capture innovation efforts. For a formal definition of R&D activities, we turn to the family of manuals developed by the organization for Economic Co-operation and Development (OECD), specifically the 2015 Frascati Manual and the 2018 Oslo Manual, where for an activity to be classified as R&D, it needs to be (i) novel, (ii) creative, (iii) uncertain in outcome, (iv) systematic, and (v) transferable or reproducible (or both). The activities reflected in our exploration indicator satisfy the first three criteria: these activities are novel and creative to the firm and its industry by construction, and the activities are positively related to uncertainty (see Section 3.5). Furthermore, we cannot rule out the possibility that our indicator captures some systematic and transferable/reproducible activities. The activities captured in the exploration indicator, therefore, are expected to overlap with some R&D activities.

Existing academic literature further classifies innovation activities into exploratory and exploitative. Exploratory activities, which involve

[15] For example, the word "cloud" has a different meaning in agriculture than it has in computing.

[16] Young (2009) distinguishes among three different types of diffusion mechanisms – contagion, social influence, and social learning.

[17] Because around seven percent of all new nouns in our sample have no prior references, the 2.5$^{\text{th}}$ percentile of $NF_{i,t}(\omega)$ is not well-defined. Instead, we use the 16$^{\text{th}}$ percentile of $NF_{i,t}(\omega)$ to define explorations.

Table 1
Distributional characteristics of the exploration indicator.

	Obs.	% firm-years with explorativ words	Mean	Standard deviation	First quartile	Median	Third quartile
Industry Group (from most to least explorative)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pharma, Biotech & Life Sciences	5,642	97.7	12.9	10.1	5.0	10.0	18.0
IT Software & Services	7,303	93.7	9.6	9.4	3.0	7.0	13.0
Communications & Entertainment	3,936	93.6	10.1	10.4	3.0	7.0	14.0
Health Care Equipment & Services	6,480	93.4	8.3	8.4	3.0	6.0	11.0
Energy	4,166	91.0	8.2	8.8	2.0	5.0	11.0
IT Hardware & Semiconductors	8,821	91.0	7.8	8.2	2.0	5.0	11.0
Utilities	2,179	90.9	8.6	10.1	2.0	5.0	11.0
Consumer Services	3,232	90.2	7.5	8.1	2.0	5.0	10.0
Consumer Goods	4,421	88.0	6.4	7.8	1.0	4.0	8.0
Transportation	1,529	87.8	6.9	8.1	2.0	4.0	9.0
Commercial & Professional Services	3,332	87.2	6.3	7.4	1.0	4.0	8.0
Materials	3,601	86.5	6.2	7.9	1.0	4.0	8.0
Retailing	3,844	86.1	5.9	7.4	1.0	3.0	7.0
Capital Goods	6,694	84.9	5.8	7.1	1.0	3.0	8.0
Financials	15,584	84.6	5.3	6.9	1.0	3.0	7.0
Consumer Staples	3,490	84.0	6.1	7.7	1.0	3.0	8.0
Real Estate	3,921	82.5	5.8	7.7	1.0	3.0	7.0
Overall	88,175	88.9	7.4	8.5	2.0	5.0	10.0

extensive search that seeks new knowledge, typically precede exploitative activities, which involve intensive search that builds on the firm's existing knowledge (see, for example, March, 1991; Quintana-García and Benavides-Velasco, 2008; Manso, 2011). We construct our language-based exploration indicator based on the idea that exploration, which aims to produce new knowledge, requires firms to introduce new words or new uses of existing words in their vocabulary. In contrast to exploration, exploitation builds on the knowledge acquired through prior exploratory activities and, therefore, is less likely to require new vocabulary. While the exploration indicator may capture some exploitative activities of the firm, it is mainly positioned to capture exploratory efforts. In comparison with the exploration indicator, R&D includes both exploratory and exploitative innovation activities. The 'research' part of R&D comprises basic and applied research activities, both of which involve the acquisition of new knowledge (OECD, 2015). The 'development' part of R&D involves activities that may build on the firm's existing knowledge and are directed towards producing new products or processes or towards improving existing products or processes.[18]

3. Validation of the exploration indicator

In this section, we discuss the distributional characteristics of the exploration indicator and present our main validation tests.

3.1. Properties of the indicator

Table 1 reports summary statistics of the exploration indicator for the overall sample and in each one of 17 industry groups. Column (2) reports the percentage of firm-year observations with a positive value of the exploration indicator, sorted from highest to lowest. The last row reports statistics for the overall sample. The indicator classifies Pharma, Biotech & Life Sciences as the most explorative industry group (97.7 firm-years with at least one explorative word) and Real Estate as the least explorative industry group (82.5 firm-years with at least one explorative word).

Columns (3) to (7) summarize the exploration indicator. We observe that firms introduce around 7.4 explorative words per year, on average, with a median of around 5.0 explorative words (last row of Table 1). The introduction of explorative words varies significantly within sample firms, with the first quartile at 2.0 explorative words and the third quartile at 10.0 explorative words per year. The table shows a meaningful variation in explorative words and exploratory activities across industries. For example, industries such as Pharma, Biotech & Life Sciences introduce 12.9 explorative words per year, whereas the Finance industry introduces only 5.3 explorative words per year.

The patterns of the exploration indicator across industry groups are informative. For example, they reveal that our variable captures potential exploration over the full spectrum of industrial activities in the U. S., where even the least explorative industries exhibit non-trivial indications of exploration. Moreover, our variable captures the across-

Table 2
Top-3 most explorative firms by industry group.

Industry Group (from most to least explorative)	*Company #1*	*Company #2*	*Company #3*
Pharma, Biotech & Life Sciences	*PerkinElmer*	*Vertex Pharmaceuticals*	*Ionis Pharmaceuticals*
IT Software & Services	*MicroStrategy*	*American Software*	*Adobe*
Communications & Entertainment	*CBS*	*Starz*	*Time Warner*
Health Care Equipment & Services	*Invacare*	*Antares Pharma*	*Quest Diagnostics*
Energy	*Chevron*	*ConocoPhillips*	*Anadarko Petroleum*
IT Hardware & Semiconductors	*Trimble*	*VIAVI Solutions*	*Digital Angel*
Utilities	*AES*	*Entergy*	*NRG Energy*
Consumer Services	*Graham Holdings*	*Adtalem Global Education*	*WMS Industries*
Consumer Goods	*Hasbro*	*Mattel*	*Tenneco*
Transportation	*FedEx*	*Avis Budget Group*	*Matson*
Commercial & Professional Services	*Exponent*	*Tetra Tech*	*Odyssey Marine Exploration*
Materials	*Alcoa*	*DuPont*	*Freeport-McMoRan*
Retailing	*Altaba*	*Pier 1 Imports*	*Hot Topic*
Capital Goods	*L3Harris Technologies*	*FuelCell Energy*	*Raytheon*
Financials	*American Express*	*Berkshire Hathaway*	*Loews*
Consumer Staples	*Lorillard*	*Spectrum Brands*	*Nu Skin Enterprises*
Real Estate	*Jones Lang LaSalle*	*GEO Group*	*St. Joe Company*

[18] See also Czarnitzki, Kraft and Thorwarth (2009), Czarnitzki, Hottenrott and Thorwarth (2011) and Barge-Gil and López (2014a, 2014b). Distinguishing research from development is particularly important given that research activities, which tend to precede development activities, are found to be a stronger driver for firm productivity (Griliches, 1986; Czarnitzki, Kraft and Thorwarth, 2009).

Table 3
Distribution of explorative words by category.

Category	Percent by Category Product	Process	Market	Supply	Organization
Pharma, Biotech & Life Sciences	34.78	15.65	21.45	7.54	20.58
IT Software & Services	46.40	5.98	37.26	2.28	8.08
Communications & Entertainment	71.67	10.34	10.10	1.48	6.40
Health Care Equipment & Services	34.13	24.34	22.67	4.30	14.56
Energy	11.71	37.84	18.92	5.41	26.13
IT Hardware & Semiconductors	67.59	13.89	13.89	0.00	4.63
Utilities	7.00	19.33	27.00	14.00	32.67
Consumer Services	31.09	28.57	30.25	0.84	9.24
Consumer Goods	57.58	18.18	16.67	1.52	6.06
Transportation	18.75	31.25	25.00	3.57	21.43
Commercial & Professional Services	52.32	8.61	31.79	0.00	7.28
Materials	33.33	17.44	22.05	4.62	22.56
Retailing	14.78	36.52	11.30	8.70	28.70
Capital Goods	34.68	13.29	32.95	2.31	16.76
Financials	10.85	20.93	34.11	3.88	30.23
Consumer Staples	42.11	14.74	24.21	5.26	13.68
Real Estate	7.69	10.00	58.46	3.08	20.77
Overall	36.10	17.41	25.70	4.38	16.42

industry variation in potential exploratory activities in an anticipated manner.

Next, we present a list of the three most explorative firms according to our exploration indicator for each one of the 17 industry groups. To calculate exploration at the firm level, we aggregate the exploration indicator for each firm over all available years. Table 2 shows the resulting list. Most of the firms that make the list are well-known names. For example, in Consumer Goods, Hasbro and Mattel rank highest based on our indicator of exploratory activities. In the energy space, Chevron and ConocoPhillips top the list while American Express and Berkshire Hathaway rank as the most explorative firms in the Financials industry group.[19]

To provide further insights into the nature and diversity of exploratory activities captured by our indicator, we look at the list of explorative words used by firms from the vantage point of the framework advanced by Schumpeter (1934). Schumpeter (1934) argues that innovative activities of firms are a complex phenomenon, where such activities can be broadly classified into five types: activities associated with the (i) creation of new products, (ii) creation and implementation of new processes, (iii) exploration of new markets, (iv) adoption of new supply chain structures, and (v) introduction of new organizational structures. For this analysis, we examine a sub-sample of firms and manually classify, based on context, each of the explorative words used by these firms into one of the above five categories. To construct the sub-sample, each year we rank all firms based on asset size and calculate each firm's average size percentile over the whole sample. Then, we select three firms corresponding to the first (small), second (medium), and third (large) quartile of asset size in each of the 17 industry sectors in our sample. These 51 firms introduced 3,654 explorative words over our sample period.[20] We manually read all related Form 10-K annual filings of these firms and, based on the specific context, classified each of the 3654 explorative words into one of the five categories advanced by Schumpeter (1934): product, process, market, supply, and organization.

Table 3 reports the results from this classification process. Examining the last row of the table, we see that about 36.1 percent of the explorative words are classified as words associated with exploratory activities in the product space, followed by the market and process categories with 25.7 percent and 17.4 percent respectively. As one would expect, the share of product-related explorative words is highest in product-oriented sectors such as Communications & Entertainment, IT Hardware, and Consumer Goods.[21] At the other end of the spectrum, the number of explorative words in the product category is lowest for highly regulated industries such as the Utilities and the Energy sectors.

One concern with a language-based indicator is that firms may try to attract a target audience or investors by adopting new rhetorical words or symbolisms (i.e., spin-writing), which may misrepresent or obfuscate real activities. We think that this scenario is unlikely in our setting. First, the information presented in Form 10-K annual filings is considered material, and firms and top managers bear legal responsibility for purposefully misrepresenting their activities.[22] Second, to further control for spin-writing, our exploration indicator includes only new words that have not been widely used by other firms in the industry at the time of adoption. Such words are less likely to be attention-grabbing and to be used strategically by firms. Third, in additional tests (untabulated for brevity) we examine the time variation of our exploration indicator and consider two notable disclosure events related to the possibility of spin-writing. The first event is the implementation of the SEC's plain English initiative of 1998 (see, for example, Loughran and McDonald, 2014) and the second event is the passage of the Sarbanes-Oxley (SOX) Act of 2002, which we also discuss in footnote 20. If the proposed exploration indicator captures spin-writing, then we expect to see significant declines in the indicator in the years immediately following the two events. Inconsistent with these expectations, we find that the two events in question are not followed by significant declines in the proposed exploration indicator.[23]

An alternative test on whether firms adopt new words with the intent of appearing trendy is provided by examining the subsequent adoption of new words in the market. In Table 4, we find that, on average, 62.6 percent of the explorative words are never adopted by industry peers and only 3.3 percent are adopted by ten or more peers (last row). The low subsequent adoption rates are inconsistent with firms being eager to replicate the disclosures of their peers. Instead, the reported findings are consistent with the idea that exploratory activities are risky, thus

[19] Quest Diagnostics, a top three Pharma firm in our list is at the forefront of COVID19 testing in the USA (Dukakis, 2020). This observation provides anecdotal evidence that our text-based exploration indicator is able to predict the accumulation of knowledge by firms "out-of-sample."

[20] Drejer (2004) demonstrates the applicability of Schumpeter's classification to both services and manufacturing.

[21] While a direct comparison with prior studies is difficult, it is worth noting that IT Hardware includes industries, such as instruments and electrical-electronic engineering, which also rank high in product innovations in earlier studies (see among others Pavitt, 1984 and Pavitt, Robson and Townsend, 1989).

[22] Materiality is one of the core principles of financial reporting established by the federal securities laws. In 1982, the Commission revised Rule 12b-2 to adopt the Supreme Court's definition of materiality. According to this definition, information is material if there is a substantial likelihood that a reasonable investor would consider the information important in deciding how to vote or make an investment decision. The Court further explained that information is material if there is a substantial likelihood that disclosure of the omitted fact would have been viewed by a reasonable investor as having significantly altered the "total mix" of information available. Following the passage of the Sarbanes-Oxley (SOX) Act of 2002, a company's CEO and CFO must each personally provide a certification as part of the company's annual Form 10-K filing. These certifications are required under Sections 302 and 906 of SOX. When certifying Section 302 disclosures, the principal officers are, among other things, (i) confirming they have reviewed the report and (ii) stating that, based on their knowledge, the report does not contain false or misleading statements or omit necessary material information. Section 906 of SOX puts forward criminal penalties for certifying a misleading or fraudulent financial report.

[23] The estimates from these additional tests are available upon request.

Table 4
Future adoptions of explorative words.

Industry Group(from most to least explorative)	Obs.	Number of Future Adopters per Explorative Word (percent of observations) 0	1	2	3	4	5	6	7	8	9	> 10
Pharma, Biotech & Life Sciences	50,582	58.2	15.6	7.0	4.2	2.8	2.1	1.6	1.3	1.0	0.8	5.4
IT Software & Services	42,555	60.5	15.2	6.9	4.1	2.7	1.9	1.5	1.1	0.9	0.7	4.6
Communications & Entertainment	32,301	62.2	16.3	6.9	4.0	2.6	1.8	1.3	1.0	0.7	0.6	2.6
Health Care Equipment & Services	38,453	63.0	15.0	6.5	4.1	2.6	1.8	1.3	1.1	0.8	0.6	3.2
Energy	22,468	59.0	15.6	7.4	4.5	3.0	2.4	1.4	1.3	0.9	0.8	3.8
IT Hardware & Semiconductors	46,068	64.7	14.1	6.1	3.6	2.5	1.6	1.3	0.9	0.9	0.6	3.6
Utilities	15,267	57.5	17.3	8.1	5.0	3.1	1.9	1.4	1.3	0.7	0.6	3.1
Consumer Services	18,038	61.3	15.8	7.5	4.5	2.8	1.9	1.3	1.1	0.8	0.6	2.3
Consumer Goods	23,313	67.3	15.0	6.6	3.6	2.3	1.4	1.0	0.7	0.5	0.3	1.2
Transportation	8,870	64.7	16.7	7.1	3.6	2.2	1.5	1.0	0.7	0.5	0.4	1.6
Commercial & Professional Services	18,634	69.4	14.3	6.4	3.4	2.1	1.4	0.9	0.6	0.3	0.3	1.0
Materials	20,640	70.1	15.5	5.8	3.1	1.9	1.2	0.6	0.5	0.2	0.2	0.8
Retailing	16,918	63.2	15.5	7.4	4.3	2.8	1.6	1.2	0.8	0.7	0.5	2.0
Capital Goods	31,077	67.4	14.5	6.4	3.4	2.3	1.5	1.1	0.7	0.5	0.5	1.8
Financials	29,869	54.9	14.7	7.1	4.4	3.0	2.1	1.9	1.3	1.2	1.0	8.4
Consumer Staples	19,526	68.5	16.0	6.3	3.1	2.0	1.3	0.8	0.5	0.4	0.3	0.9
Real Estate	15,828	58.2	16.4	7.8	4.6	3.2	2.2	1.5	1.3	0.9	0.7	3.2
Overall	450,407	62.6	15.3	6.8	4.0	2.6	1.8	1.3	1.0	0.8	0.6	3.3

Table 5
Relation of the exploration indicator with R&D and firm involvement with new technologies within and across industry groups.

Industry Group (from most to least explorative)	% firm-years with explorative words (1)	% withR&D (2)	% withnew tech (3)	Within-industry partial correlations of the exploration indicator and: ln(1 + $R\&D$) (4)	Num. new techs (5)
Pharma, Biotech & Life Sciences	97.7	95.5	6.9	0.199***	0.083***
IT Software & Services	93.7	73.5	63.8	0.126***	0.228***
Communications & Entertainment	93.6	21.2	46.5	0.070***	0.053***
Health Care Equipment & Services	93.4	60.8	14.1	0.243***	0.069***
Energy	91.0	12.1	14.6	0.196***	0.089***
IT Hardware & Semiconductors	91.0	89.1	48.8	0.198***	0.228***
Utilities	90.9	1.5	4.0	0.109***	0.107***
Consumer Services	90.2	9.0	10.6	0.193***	0.243***
Consumer Goods	88.0	41.9	6.1	0.192***	0.113***
Transportation	87.8	1.3	7.1	0.001	0.154***
Commercial & Professional Services	87.2	21.5	14.6	0.043**	0.120***
Materials	86.5	56.4	2.4	0.151***	0.082***
Retailing	86.1	7.1	12.9	0.210***	0.207***
Capital Goods	84.9	64.3	9.7	0.186***	0.187***
Financials	84.6	1.6	5.0	0.202***	0.258***
Consumer Staples	84.0	38.4	5.6	0.098***	0.104***
Real Estate	82.5	2.1	3.6	0.191***	0.185***
Overall	88.9	39.8	18.5	0.207***	0.149***
Across-industry group correlation with (1)		0.479	0.481		
(*p*-values)		(0.052)	(0.050)		

Note: *** $p<0.01$, ** $p<0.05$, *** $p<0.10$.

exhibiting high rates of failure and low rates of future adoption.[24]

3.2. Relation to other exploration indicators

3.2.1. R&D expenses

In this section, we study the empirical relation between our indicator and proxies, such as R&D spending, which also capture early-stage exploratory activities. Research is usually considered as the first step in the innovation process. A large literature has emphasized that spending on R&D leads to innovation and advances in technological knowledge (Aghion and Howitt, 1992 and 1998a; Akcigit and Kerr, 2018; Grossman and Helpman, 1991; Romer, 1986). As a result, if our indicator captures exploratory activities then it will be positively correlated with R&D spending.

Table 5 examines the relationship between our exploration indicator and R&D spending for the overall sample as well as across and within each of the 17 industry groups discussed in the previous section. Columns (1) and (2) of the table report the percentage of firm-years with positive exploration indicator and R&D spending, respectively. Only 39.8% of the firm years report positive R&D expenses. Compared to this, our language-based exploration variable indicates exploratory activities in 88.9% of the firm-years. The industries in Table 5 are sorted based on the number firm-years with at least one explorative word. Both this ranking and ranking based on R&D expenses sort Pharma, Biotech & Life Sciences at the top. Other R&D-intensive industries such as IT Software & Services and IT Hardware & Semiconductors are also ranked high by our variable. The similarity in ranking across industry groups is reflected in the last row of column (2), which shows that the across-industry correlation between the exploration indicator and R&D expenses is 0.479, suggesting that our language-based variable indeed captures exploratory activities of firms.

[24] For example, the low adoption rates are consistent with the high failure rates of high-tech startups. High-tech startups are typically innovation-intensive firms with very high-growth potential that raise funds from angel investors and venture capitalists. The entrepreneurial financing literature suggests that more than 75 percent of startups fail to become successful operating firms (Kerr, Nanda and Rhodes-Kropf, 2014).

Column (4) shows the within-industry partial correlation between the exploration indicator and R&D expenses, where the partial correlations control for year-fixed effects. The correlation coefficients are positive and range between 0.001 and 0.243. Moreover, for 15 out of 17 industries the coefficients are positive and significant at 0.01 level.

Reporting of R&D spending was originally formalized to capture R&D activities in the labs of manufacturing firms. While subsequently it was also adopted by the services sector, R&D spending does not fit as well for service firms and is likely to underestimate exploratory efforts for such firms.[25] This warrants a closer examination of the relation between R&D spending and our exploration indicator separately for manufacturing and for service firms. To determine whether a firm operates in the manufacturing or services sectors, we use the divisions defined by the Standard Industrial Classification system. Manufacturing firms are those firms within division D (Manufacturing). Service firms are those firms within divisions E (Transportation, Communications, Electric, Gas, and Sanitary Services), F (Wholesale Trade), G (Retail Trade), H (Finance, Insurance, and Real Estate), I (Services), and J (Public Administration). Consistent with expectations, the partial correlation between R&D spending and our exploration indicator is higher for manufacturing firms (correlation of 0.222) than for service firms (correlation of 0.180). The difference of 0.042 has a standard deviation of 0.007 and is significant at the 0.001 level.

We note that, while positively related to R&D spending, our variable also exhibits significant variation unrelated to R&D spending. The contention in this paper is that this unrelated variation is because our exploration indicator captures dimensions of the creative and innovation process that are different than those captured by R&D spending.[26] In subsequent sections, we provide evidence that our exploration indicator precedes R&D spending and in general exhibits dynamics that are substantially different than those of R&D spending. Moreover, we provide evidence that our exploration indicator is highly informative about the future accumulation of knowledge, firm risk, and firm value, even after controlling for R&D spending.

3.2.2. Gartner's innovative technologies list

For an alternative validation test, we compile a list of 204 emerging technologies between 2000 and 2016 from the Emerging Technologies Hype Cycle yearly reports by Gartner Inc.[27] Gartner is a leading IT consulting firm that conducts surveys of public and private companies, governments, academics, industry experts and analysts to create an annual qualitative ranking of notable new technologies that have the potential to become mainstream. Gartner's reports are widely used by practitioners and corporations to make strategic investment decisions (Steinert and Leifer, 2010). Gartner's technology classifications have also received recent attention in a wide range of academic disciplines, including Accounting, Computer Science, Management Information Systems, Medicine, and Technology Innovation Management (Jarvenpaa and Makinen, 2008; O'Leary, 2008; Dedehayir and Steinert, 2016; Heading, 2017; Chen and Han, 2019).

We posit that firms attempting exploration would experiment with some of these emerging technologies irrespective of whether these technologies succeed or fail later. Such experimentation would be also reflected in the language firms use to describe their business operations. As an illustration, the concept of "Internet TV" appeared on Gartner's list for the first time in the year 2009. However, the term Internet TV was introduced and discussed as early as 1998 by three firms in our sample—Echostar Communications, Boca Research Inc. (now ENER1 Inc.), and C Cube Microsystems Inc. All three companies were either introducing Internet TV services or were discussing plans to introduce the service in their 1998 annual filings. For example, while discussing the line-up of new offerings, Echostar Communications stated this: "*We are expanding our offerings to include Internet and high-speed data services… This service integrates DISH Network's digital satellite television programming with Internet TV services from WebTV.*" Another example is the "Cloud Computing" technology that appeared in Gartner's list for the first time in 2008. Callwave Inc. (now Fuzebox Software), a business communications company, commented in its 2007 filing that "*CallWave plans to use cloud computing to scale its business without incurring traditional costs that are associated with growth of operations.*"

Not all technologies on the Gartner list, however, become successful. Our variable is expected to capture a wider range of exploratory efforts of firms, even efforts that do not succeed. One example in this regard is the "Ultrawideband Signal Processing" technology that appeared in the 2003 Gartner's list. Essex Corporation (now part of Northrop Grumman), an IT Hardware & Semiconductor company, discussed this technology in their 1998 annual filing. This technology is never mentioned by another company after 2000.

To operationalize our test, we construct a variable that counts the number of new technologies from Gartner's list referenced by each firm in its annual filing each year. Column (3) of Table 5 reports the percentage of firm-years with at least one new technology on Gartner's list that is discussed in a firm's business description. The numbers show that firms in IT Software & Services, IT Hardware & Semiconductors, and Communications & Entertainment are the most likely to engage in the new technologies on Gartner's list. The across-industry correlation between the exploration indicator and Gartner's new technology variable expenses is 0.481 implying that our language-based indicator is able to capture activities of firms that explore the use of new technologies. Column (5) shows the within-industry partial correlations between exploration and Gartner new technologies variable, where the partial correlations control for year fixed effects. We note that all correlation coefficients are positive and significant at 0.01 level, ranging between 0.051 and 0.258. These findings further confirm that our exploration indicator captures significant variations in the exploratory efforts of firms.

3.3. The dynamic relations between exploration, patents, trademarks, and R&D

In this section, we investigate the dynamic relations among our exploration indicator, firm R&D spending, and firm patenting and trademarking activities. When measuring patenting and trademarking activities, we distinguish between application dates and granting dates, which allows us to examine in more detail how exploratory activities, as captured by our proposed indicator, are related to the timing of the patenting and trademarking processes.

3.3.1. Exploration and patents

Patents are commonly seen as an indicator of innovative output (Griliches, 1990; Bloom and Van Reenen, 2002; Hall et al., 2005). They provide inventors with limited-term rights to the innovation and preclude others from generating profits from the innovation. Patents, however, vary substantially in terms of their quality, technological and economic importance, which makes the distribution of their value highly skewed (Hall et al., 2005). Consequently, raw patent counts do not account for the heterogeneity of innovation quality, i.e., the value of the "output" associated with a particular patent (Griliches et al., 1986). Jaffe and Trajtenberg (2002) note that patent citations represent the

[25] See also Kleinknecht, Van Montfort and Brouwer (2002) and Oslo Manual (OECD, 2018) on the collection of innovation indicators.

[26] As a related note, the reporting of R&D expenditures in financial statements is affected by tax considerations (Lerner and Seru, 2021), while our exploration indicator is not directly affected by such considerations.

[27] The list is available from the authors upon request. Gartner Inc. is an S&P500 company that provides research and advisory services to governments and small and large enterprises, including to 77% of the 500 largest corporations in the world. For more information, see Understanding Gartner's Hype Cycles (available at https://www.gartner.com/en/documents/3887767/understanding-gartner-s-hype-cycles).

Table 6
The dynamics of the exploration indicator, R&D, and patents.

Dependent variable:	ln(1 + Citation-weighted patent applications $_t$)	ln(1 + Citation-weighted patent grants $_t$)	ln(1 + R&D $_t$)	Exploration indicator $_t$
	(1)	(2)	(3)	(4)
Exploration indicator $_{t-1}$	0.247***	0.046	0.128***	
	(0.037)	(0.035)	(0.028)	
Exploration indicator $_{t-2}$	0.169***	0.073**	0.090***	
	(0.035)	(0.034)	(0.027)	
Exploration indicator $_{t-3}$	0.187***	0.133***	0.054**	
	(0.033)	(0.032)	(0.026)	
ln(1 + R&D $_{t-1}$)	2.771***	5.585***		– 10.526
	(0.910)	(0.869)		(11.063)
ln(1 + R&D $_{t-2}$)	– 0.675	3.544***		9.473
	(1.105)	(1.055)		(13.427)
ln(1 + R&D $_{t-3}$)	– 11.314***	2.720***		– 48.722***
	(0.849)	(0.810)		(10.355)
ln(1 + Citation-weighted patent applications $_{t-1}$)			– 0.699	41.000***
			(0.428)	(6.736)
ln(1 + Citation-weighted patent applications $_{t-2}$)			0.609	17.971**
			(0.452)	(7.101)
ln(1 + Citation-weighted patent applications $_{t-3}$)			0.100	23.362***
			(0.456)	(7.173)
ln(1 + Citation-weighted patent grants $_{t-1}$)			3.550***	– 3.357
			(0.447)	(7.036)
ln(1 + Citation-weighted patent grants $_{t-2}$)			2.185***	4.259
			(0.417)	(6.558)
ln(1 + Citation-weighted patent grants $_{t-3}$)			2.280***	– 18.161***
			(0.388)	(6.105)
ln(Assets $_t$)	7.992***	7.901***	23.839***	124.515***
	(0.509)	(0.486)	(0.387)	(6.189)
Firm- & year-fixed effects	Yes	Yes	Yes	Yes
Observations	53,135	53,135	53,135	53,135
Adjusted R-squared (%)	77.98	84.67	96.22	41.27

Note: Standard errors in parentheses *** $p<0.01$, ** $p<0.05$, *** $p<0.10$.

cumulative process of knowledge development that future inventors use and build on. To be consistent with recent literature (Hall et al., 2005; Kogan et al., 2017), we use citation-weighted patent counts as a knowledge metric.

We create two citation-weighted patent count variables, one based on the patent application date and the other based on the patent grant date. When constructing these citation-weighted variables, we need to consider that a patent can generate citations only after it has been granted so that, all else equal, a patent with an earlier grant date will have accumulated more citations than a patent with a later grant date. To control for this aspect of the patent data, we define "per-year citations" as the total number of citations a patent has received by December 31, 2019, divided by the number of years from the patent grant date until December 31, 2019. Citation-weighted patent applications for firm *i* in year *t* are then calculated as the sum, over all patent applications filed by firm *i* in year *t*, of the number of per-year citations generated by each patent. Citation-weighted patent grants are calculated in a similar manner, using patent grants in a given firm-year rather than patent applications.

While we expect an overall positive relation between our exploration indicator and both patent applications and grants, the dynamics of these relations are potentially complex. Because developing a valuable patent requires exploratory efforts, we expect our language-based exploration indicator to predict future patent applications. However, patents do not necessarily constitute a "final" product, so that we expect exploratory efforts, as captured by the emergence of new terminology, to continue after patent applications and possibly even after patent grants.

Interpreting R&D spending also poses a challenge: part of R&D reflects research (the "R" of R&D) and part of it reflects development (the "D" of R&D). Because development likely follows exploratory efforts, we expect our exploration indicator to be able to also predict R&D spending. We have no ex ante expectations of how R&D spending should be related to subsequent exploratory efforts, as measured by our exploration indicator. On the one hand, it is possible that an increase in R&D spending may spur future exploratory efforts. Moreover, the firm may coin new terminology only after it has already performed some R&D activities. On the other hand, however, existing evidence suggests that R&D spending follows waves, where periods of relatively high R&D are followed by periods of relatively low R&D, as firms ramp up manufacturing efforts (Brown et al., 2009). If exploratory activities also correlate with this pattern, it is possible that R&D spending may be negatively related to future exploratory activities. When forming expectations about the timing of R&D spending in relation to patent applications and grants, we again consider that a new patent application may be related to research, which was carried out before the application, while a new grant may be related to development, which could be carried out during and after the patent is granted. Thus, we expect R&D spending to precede patent applications but also to precede and even possibly follow patent grants.

We start our analysis by examining how patent applications are related to past exploratory activities, as measured by our exploration indicator. In Table 6, model (1), we use citation-weighted patent applications as the dependent variable. The explanatory variable of main interest is the language-based exploration indicator, measured as of years *t*-1, *t*-2, and *t*-3.[28] As discussed above, R&D is also likely to be related to patenting activities, so that we control for past R&D spending. Furthermore, we control for firm asset size and include firm- and year-fixed effects. The inclusion of firm-fixed effects allows us to take into account latent variation in patenting activity that is specific to each firm. For the sake of readability, the reported coefficient estimates and their standard errors are multiplied by 100.

Looking at the estimates in column (1) of Table 6, we find that the introduction of explorative words in years *t*-1, *t*-2, and *t*-3 is positively and significantly related to citation-weighted patent applications by a

[28] We present and discuss estimates based on three lags, mainly because the gap between patent applications and patent grants is around three years, on average. However, we have verified that our findings are not sensitive to including fewer than three or more than three lags.

Table 7
The dynamics of the exploration indicator, R&D, and trademarks.

Dependent variable:	ln(1 + Trademark applications $_t$)	ln(1 + Trademark grants $_t$)	ln(1 + R&D $_t$)	Exploration indicator $_t$
	(1)	(2)	(3)	(4)
Exploration indicator $_{t\text{-}1}$	0.188***	0.137***	0.125***	
	(0.042)	(0.036)	(0.028)	
Exploration indicator $_{t\text{-}2}$	0.088**	0.191***	0.091***	
	(0.040)	(0.035)	(0.027)	
Exploration indicator $_{t\text{-}3}$	– 0.008	0.133***	0.057**	
	(0.038)	(0.033)	(0.026)	
ln(1 + R&D $_{t\text{-}1}$)	2.048**	– 0.081		– 7.152
	(1.036)	(0.899)		(11.072)
ln(1 + R&D $_{t\text{-}2}$)	– 1.177	1.547		10.447
	(1.257)	(1.092)		(13.439)
ln(1 + R&D $_{t\text{-}3}$)	– 4.217***	– 0.872		– 53.544***
	(0.965)	(0.838)		(10.325)
ln(1 + Trademark applications $_{t\text{-}1}$)			0.136	13.807***
			(0.315)	(4.946)
ln(1 + Trademark applications $_{t\text{-}2}$)			0.564	12.423**
			(0.352)	(5.516)
ln(1 + Trademark applications $_{t\text{-}3}$)			– 0.288	5.971
			(0.369)	(5.792)
ln(1 + Trademark grants $_{t\text{-}1}$)			– 0.650	– 16.506**
			(0.437)	(6.861)
ln(1 + Trademark grants $_{t\text{-}2}$)			0.727*	3.768
			(0.410)	(6.442)
ln(1 + Trademark grants $_{t\text{-}3}$)			0.791**	– 9.034
			(0.369)	(5.792)
ln(Assets $_t$)	9.502***	6.421***	24.445***	129.695***
	(0.579)	(0.503)	(0.387)	(6.195)
Firm- & year-fixed effects	Yes	Yes	Yes	Yes
Observations	53,135	53,135	53,135	53,135
Adjusted R-squared (%)	64.79	64.77	96.20	41.13

Note: Standard errors in parentheses *** $p<0.01$, ** $p<0.05$, *** $p<0.10$.

firm in year *t*. These findings collaborate the idea that exploration does not only relate to patenting activity but also is able to predict such activity. To facilitate the interpretation of our estimates, Table A.1 in Appendix A reports the means and standard deviations of the different variables used in the analysis. Based on the reported estimates, we find that a one standard deviation increase in the number of explorative words in year *t*-1 is associated with an increase in citation-weighted patent applications by around 0.021 standard deviations (or (0.247/100) × (8.4723/0.9917)). Cumulatively, a one standard deviation increase in the exploration indicator in each of the last three years is associated with an increase in citation-weighted patent applications by around 0.052 standard deviations.

R&D spending in year *t*-1 also is positively related to patent applications in year *t*, where a one standard deviation increase in R&D spending in year *t*-1 is related to an increase in patent applications by around 0.050 standard deviations (or (2.771/100) × (1.7812/0.9917)). However, unlike our exploration indicator, for which all three lags are related to future patent applications, R&D spending in year *t*-3 is negatively related to patent applications in year *t*. These findings suggest that R&D spending increases close to patent applications while exploratory activities, as measured by our indicator, tend to consistently increase in the several years prior to patent applications.

In Table 6, model (2), we perform a similar analysis but now use citation-weighted patent grants in year *t* as the dependent variable. Now we find that the coefficient on our exploration indicator as of *t*-1, although positive, is insignificant. But we again find positive and significant coefficients on the exploration indicator variables for *t*-2 and *t*-3. Cumulatively, a one standard deviation increase in the exploration indicator in each of the last three years is associated with an increase in citation-weighted patent grants by around 0.019 standard deviations in the current year. The difference in estimates between patent applications and patent grants suggests that, when compared to the more traditional indicator of R&D spending, our proposed exploration indicator tends to pick up relatively early-stage exploratory activities. Consistent with this interpretation, in model (2), we find that all three lags of R&D spending are positively related to patent grants.

To examine the intertemporal relation between our exploration indicator and R&D spending, in model (3) of Table 6, we estimate a similar model where the dependent variable now is R&D spending as of year *t*. As the explanatory variable of main interest, we again use our exploration indicator as of years *t*-1, *t*-2, and *t*-3. We also include three lags of the patent applications and patent grants variables, which allows us to also examine how R&D spending relates to these key patenting events. We find that our exploration indicator is positively related to future R&D spending, for all three lags. For an economic interpretation, we find that a one standard deviation increase in the exploration indicator in each of the last three years is associated with a cumulative increase in R&D spending by 0.013 standard deviations. Looking at the patenting variables, we find that R&D spending is positively related to past patent grants, which is consistent with the idea that the development portion of R&D may follow patent grants.[29]

We also estimate a model where R&D is the dependent variable and as explanatory variables we use not only the past three lags but also the contemporaneous value of our exploration indicator. The estimates, reported in Table A.3 of Appendix A, show that, after controlling for past exploration, R&D is not significantly related to the contemporaneous value of the exploration indicator. These findings confirm that the proposed indicator indeed tends to lead R&D spending. In Table A.3, we further examine the dynamic relation between our exploration indicator and R&D spending for manufacturing firms and for service firms. We find that our exploration indicator predicts R&D spending for both

[29] Interestingly, we find that past patent applications do not predict R&D spending. Because the average gap between patent application and grant is around three years, it is possible that lagged patent grants subsume the importance of lagged patent applications. To examine this possibility, we estimate an alternative model (not tabulated), in which we include only two lags of each of the explanatory variables. We now find that lagged patent applications indeed predict R&D spending. The estimates are available upon request.

manufacturing and service firms, where the effect of past exploration on R&D is generally higher for manufacturing firms.

The last model of Table 6 examines how our exploration indicator at time *t* is related to past R&D spending and patenting activities. We find that R&D spending in the past two years is not significantly related to our exploration indication. However, we also find that R&D spending in year *t*-3 is negatively related to the exploration variable in year t. These findings are consistent with the idea that exploration declines after periods of relative intensive investment in research and development activities, possibly because then firms may expend resources on ramping up manufacturing.[30] In regard to past patenting activities, we find that citation-weighted patent applications in years *t*-1, *t*-2, and *t*-3 are positively related to the language-based exploration indicator in year *t*. In contrast, the exploration indicator in year *t* is not related to citation-weighted patent grants in years *t*-1 and *t*-2 and is negatively related to patent grants in year *t*-3. These findings are consistent with the expectation that exploratory activities pick up after patent applications, but that by the time patents are granted, exploratory activities within the firm may have stabilized. The longer-term negative relation between past patent grants and current exploratory indications mirrors our finding pertaining to the relation of past R&D spending and current exploration.

The overall findings presented in this section are consistent with our expectations and more generally with the notion that the proposed exploration indicator captures early-stage exploratory activities. In particular, we find that the exploration indicator is positively related to future R&D spending and patenting activities. The exploration indicator also increases in the years immediately following patent applications. However, the findings suggest that the exploration indicator stabilizes after patent grants and R&D spending, and may eventually even decline, a decline that could be due to firms re-focusing on exploitation rather than exploration.

3.3.2. Exploration and trademarks

We employ trademark activity as a second indicator of knowledge accumulation within the firm (Mendonça et al., 2004; Dinlersoz et al., 2019; Faurel et al., 2021).[31] As in the patent-based analysis, we construct two trademark indicators: the first is the number of trademark applications filed by a firm in year *t* and the second is the number of trademark registrations granted to a firm in year *t*. Unlike patents, trademarks do not have a citation system so here we use raw trademark counts as indicators of knowledge accumulation.

Our trademark-based analysis mirrors the patent-based analysis in the previous section. We again estimate different models that examine the dynamic relations of our exploration indicator with trademark applications, trademark grants, and R&D expenditures. We again control for firm asset size and include firm- and year-fixed effects in all regressions.

The findings, presented in Table 7, are broadly analogous to the findings presented in the previous section. For example, we find that the exploration indicator is positively related to both future trademark applications (model (1)) and future trademark counts (model (2)). In model (1), we find that our exploration indicator in years *t*-1 and *t*-2 is positively related to trademark applications in year *t*, where the two coefficients are significant at the 0.01 and 0.05 levels, respectively. In terms of economic significance, a one standard deviation increase in the exploration indicator in each of the last three years is associated with an increase in trademark applications in year *t* by around 0.026 standard deviations. In model (2), we find that all three lags of the exploration indicator are positively related to trademark grants in year *t*, where all three coefficients are significant at the 0.01 level. Interpreting the estimates, a one standard deviation increase in the exploration indicator in each of the last three years is associated with an increase in trademark grants in year *t* by around 0.052 standard deviations. Confirming the results reported in Table 6, in model (3) of Table 7, we again find that our exploration indicator in years *t*-1, *t*-2, and *t*-3 is positively related to R&D spending in year *t*.

In Table 7, model (4), we examine how the exploration indicator in year *t* is related to trademark applications and trademark grants in year *t*-1, *t*-2, and *t*-3. The results show that trademark applications in years *t*-1 and *t*-2 are positively related to the exploration indicator in year *t*, with significance at the 0.01 and 0.5 levels, respectively. In contrast, trademark grants in year *t*-1 are negatively related to the exploration indicator in year *t*, where the coefficient is significant at the 0.05 level.

Examining the dynamic relations of R&D spending with trademark applications and grants across models (1) through (3), we find several results that are analogous to our findings related to patents. First, in model (1), we find that R&D spending in year *t*-1 is positively related to trademark applications in year *t* but that R&D spending in year *t*-3 is negatively related to trademark applications in year *t*. Second, in model (3), we find that R&D spending is not related to past trademark applications. Third, also in model (3), we find that R&D spending is positively related to past trademark grants. However, unlike patent grants, which are positively related to past R&D spending, in model (2) we find that trademark grants are unrelated to past R&D spending.

In sum, as far as our exploration indicator is concerned, the main takeaways from the trademark-based analysis in this section are similar to the main takeaways from the patent-based analysis in the previous section. Specifically, we find that (i) the exploration indicator is positively related to future trademark activities, (ii) the exploration indicator tends to increase in the years immediately following trademark applications, and (iii) the exploration indicator tends to stabilize or even decrease after trademark grants.

3.4. Exploration along the firm life-cycle

In this section, we study how exploration changes as firms mature. Because there are strong economic arguments on how firm exploration should change over time, and because the passage of time is exogenous, these tests allow us to further validate our indicator. There are at least three economic reasons to expect that firm's propensity to experiment declines with age. First, the incentives for incumbent firms to look for new products and services are relatively weak because such firms already enjoy some economic advantages in the marketplace (Hobijn and Jovanovic, 2001). In contrast, new players have no existing market share so in order to compete they need to introduce new products, better production technologies, or more efficient organizational structures or need to develop new supply sources or new markets. These effects could be further strengthened by the tendency of firms to lose important human capital over time. Inventors often work in teams and because departures of some inventors from the team become more likely when the firm ages, such departures could adversely affect the innovation of older firms (Jaravel et al., 2018).

Second, as the firm ages, existing agency problems between managers and outside investors are expected to exacerbate. For example, Helwege et al. (2007) show that firm equity ownership becomes gradually more diffused over time and more dispersed owners could find it more difficult to monitor corporate insiders. Consistent with this idea, Ang et al. (2000) find that agency costs increase when the firm is managed by outside rather than inside managers, when manager ownership declines, and when the number of non-manager shareholders increases. Along these lines,Martimort, (1999) further argues that as the firm ages, its insiders can more easily collude against its owners. An increase in firm agency costs over time could naturally lead to a decline

[30] These findings are broadly consistent with the evidence presented in Geroski, Van Reenen and Walters (1997), who use two different datasets (one in the U.S. and one in the U.K.) and find that very few innovative firms are persistently innovative.

[31] As is common in academic studies, we use the term "trademarks" to refer to both trademarks and servicemarks.

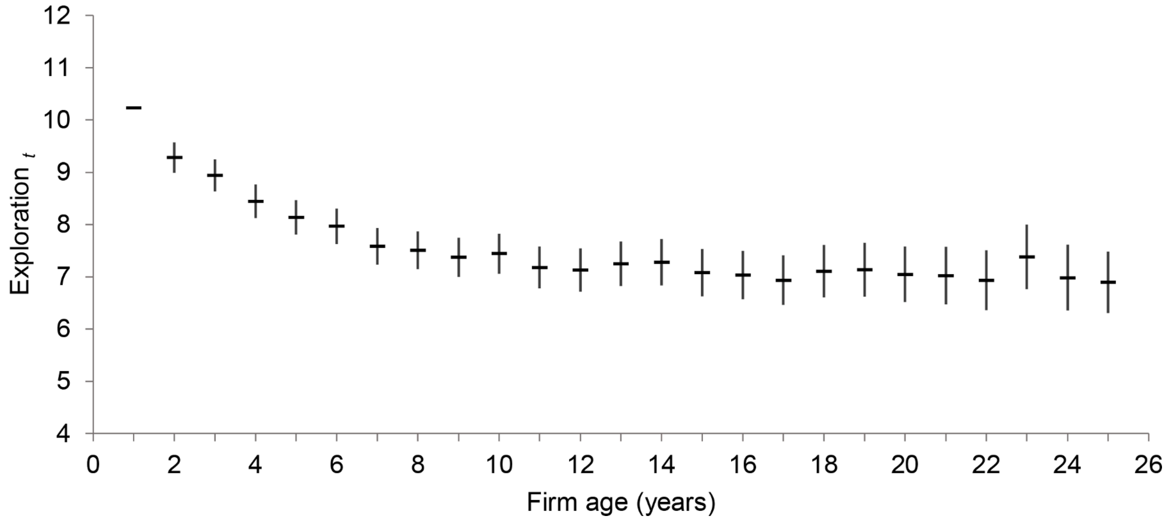

Fig. 2. Number of explorative words along the firm life cycle.

Table 8
The exploration indicator and equity risk.

Dependent variable:	Beta $_t$ (1)	(2)	Idiosyncratic volatility $_t$ (%) (3)	(4)
Exploration indicator $_{t-1}$	0.618***	0.503***	1.817***	1.696***
	(0.045)	(0.055)	(0.161)	(0.194)
Exploration indicator $_{t-2}$		0.389***		1.498***
		(0.053)		(0.187)
Exploration indicator $_{t-3}$		0.350***		1.050***
		(0.050)		(0.176)
ln(Assets $_{t-1}$)	5.392***	6.208***	– 65.510***	– 70.094***
	(0.576)	(0.706)	(2.073)	(2.486)
Firm- & year-fixed effects	Yes	Yes	Yes	Yes
Observations	86,432	62,113	86,303	62,024
Adjusted R-squared (%)	29.34	29.66	61.51	60.46

Note: Standard errors in parentheses *** $p<0.01$, ** $p<0.05$, *** $p<0.10$.

in effort-intensive yet hard to measure exploratory activities (Holmström, 1989; Francis and Smith, 1995). Third, time could position a firm on the path of an extreme or suboptimal specialization (Grossman and Shapiro 1982). In this case, a more generalist firm could find it easier to absorb negative market shocks (e.g., Dimitrov and Tice, 2006). The evidence on organizational inertia in economics is also complemented by evidence in sociology. It is well documented in the sociology (Hannan and Freeman, 1984) and economics (Nelson and Winter, 1982) literature that organizational capabilities are difficult to create and adjust.

Taking the wide range of arguments and findings in their totality, our firm exploration indicator should decline as the firm ages. To test the validity of this prediction, we estimate the following model:

$$Exploration_{it} = \alpha + \beta \ln(Assets_{it}) + \mu_i + \mu_t + \mu_k + u_{it}, \quad (2)$$

where the dependent variable is our exploration indicator and the explanatory variables include asset size and firm-fixed effects (μ_i), year-fixed effects (μ_t), and age-fixed effects (μ_k). Firm age is defined relative to a firm's first listing in CRSP and for firm ages 25 or older, we estimate one fixed effect. Fig. 2 plots the estimated firm age fixed effects with the corresponding 95% confidence intervals. The base group contains one-year-old firms and the average number of explorative words introduced by firms in this group is 10.2.

We observe that the number of explorative words that a firm introduces declines with age. The effect is the strongest over the first decade after an IPO, during which time our firm exploration indicator declines by more than 20%. The exploration indicator then tends to stabilize at around 7.5 to 7 explorative words, roughly 25% to 30% below its level at the IPO stage. These findings complement Bernstein (2015), who documents a decline in corporate innovative activity after firms go public; here we show that exploration, in general, declines in the years after the IPO.

3.5. Exploration and risk

Experimentation with new products and business practices are associated with increased uncertainty and information asymmetry. Further, March (1991) points out that compared with the returns from "exploitation of old certainties," the returns from "exploration of new possibilities" are more remote in time and less certain. Therefore, a valid exploration indicator is expected to correlate positively with risk. Since the uncertainty accompanying new projects is generally reflected in stock prices, we focus our analysis on measures of firm equity risk (Carpenter and Petersen, 2002; Pastor and Veronesi, 2009; Gu, 2016).

Table 9
The stock of exploration and firm value.

Dependent variable: ln(Market-to-book as of the end of year t)	(1)	(2)	(3)	(4)
ln(1 + Exploration stock $_t$ / Assets $_t$)	0.200***	0.161***	0.261***	0.108***
	(0.017)	(0.020)	(0.009)	(0.011)
ln(1 + R&D stock $_t$ / Assets $_t$)		0.090***		– 0.209***
		(0.025)		(0.014)
Positive R&D stock dummy $_t$		0.119***		0.044***
		(0.014)		(0.014)
ln(1 + Patent applications stock $_t$ / Assets $_t$)		0.554***		0.425***
		(0.042)		(0.022)
Firm has patent applications stock dummy $_t$		0.020*		– 0.029***
		(0.011)		(0.010)
ln(Assets $_t$)		0.017***		– 0.096***
		(0.003)		(0.003)
Industry- & year-fixed effects	Yes	Yes	No	No
Firm- & year-fixed effects	No	No	Yes	Yes
Observations	88,175	88,175	88,175	88,175
Adjusted R-squared (%)	28.44	31.16	64.02	64.79

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, *** p<0.10.

More specifically, we follow Fama and French (1992) and estimate the following equation to measure firm equity betas and idiosyncratic volatilities:

$$R_{i,w} = \alpha_i + \beta_1 R_{M,w} + \beta_2 R_{M,w-1} + u_{i,w}. \tag{3}$$

The dependent variable in Eq. (3) is the return of stock i for week w. As explanatory variables, we use the return of the CRSP equally-weighted portfolio for week w and week w-1. The above model is estimated separately for each firm-year in our sample. The final estimate of firm systematic risk is its beta, calculated as $\beta_1 + \beta_2$, which takes into account delayed price responses.

The first two columns of Table 8 report estimates from regressions that examine the link between the firm's market beta in year t, estimated from Eq. (3), and the language-based exploration indicator in years t-1, t-2, and t-3. The coefficient estimates and their standard errors are again multiplied by 100. We find a significantly positive relationship between our exploration indicator and beta. Column (1) shows that a one standard deviation increase in the exploration indicator in year t-1 is associated with a 0.052 (0.618 × 8.4723/100) increase in the firm's market beta. Examining the estimates from column (2), the cumulative effect of a one standard deviation increase in the number of explorative words introduced by a firm in each of the past three years is associated with a 0.105 increase in the firm's market beta in year t. To evaluate the economic significance of this estimate, we consider the value of a perpetuity where the relevant risk-free rate equals 3.00 percent and the market risk premium equals 4.50 percent. An increase in the beta of such a perpetuity from 1.00 to 1.10 will increase its discount rate from 7.50 percent to 7.95 percent, reducing the present value of the perpetuity from 13.33 to 12.58, or by around 5.66 percent.

In the last two columns of Table 8, we evaluate the relation between stock idiosyncratic volatility (the standard deviation of $u_{i,w}$ from Eq. (3) above) at time t and our exploration indicator over the past three years. Column (3) shows that a one standard deviation increase in the exploration indicator increases firm idiosyncratic volatility by 0.154 (1.817 × 8.4723/100) percentage points. Examining the estimates in column (4), a one standard deviation increase in exploration in each of the past three years is linked to an increase in idiosyncratic volatility by 0.360 percentage points.

The results in Table 8 underscore the point that engaging in exploratory activities is risky. One of the contributing factors to this result should be that only a small fraction of all exploratory activities lead to future value-generating opportunities.

4. The value of exploration

The wide range of findings presented in the previous section provides strong evidence that our proposed exploration indicator indeed captures relevant variation in the exploratory activities of firms. In this section, we examine an important question that naturally arises: Does exploration add economic value? Lucas Jr and Moll (2014), for instance, argue that the level of production and real growth in an economy is jointly determined by current knowledge and knowledge accumulation via exploration. Studies on the impact of patenting on firm value agree that increases in inventive activities lead to an increase in firm market value (Pakes, 1985; Bloom and Van Reenen, 2002). For example, Hall et al. (2005) show that one additional patent citation boosts the market value of a firm by three percent. In a similar line of research, Simeth and Cincera (2016) find a positive link between a company publishing articles in scientific journals and its market value. However, not all exploratory activities may result in an increase in value. For example, Belenzon and Patacconi (2013) show that while high-quality patents increase firm value, lower-quality patents have no discernable impact on firm performance.

For our analysis, we follow Hall et al. (2005) and construct stock variables for exploration as well as for patents and R&D. The stock value of any variable X for firm i in year t is constructed using the general formula:

$$Stock of X_{i,t} = \sum_{m=0}^{M} 0.85^m X_{i,t-m}. \tag{4}$$

In Eq. (4), M equals the number of historic years with available information about X for firm i up to and including year t. The formula effectively discounts the historic values of X by 15 percent per year and aggregates the discounted values up to and including year t. To calculate exploration stock, R&D stock, and patent stock we use the values of our exploration indicator, the values of R&D expenditures, and the values of citation-weighted patent applications, respectively.[32]

Table 9 reports results from regressions where the dependent variable is the market-to-book of assets (Q) for each firm, measured as of the end of year t. We calculate Q for the end of year t as the book value of total assets plus the market value of equity minus book value of equity minus deferred taxes, all divided by total assets. We estimate the following model:

[32] Section 3.3.1 outlines the construction of citation-weighted patent applications.

$$\ln(Q_{it}) = \beta_1 \ln\left(1+\frac{EXP^*_{it}}{A_{it}}\right) + \beta_2 \ln\left(1+\frac{RD^*_{it}}{A_{it}}\right) + \beta_3 D_1^{RD^*_{it}>0} + \beta_4 \ln\left(1+\frac{PAT^*_{it}}{A_{it}}\right) + \beta_5 D_2^{PAT^*_{it}>0} + \beta_6 \ln(A_{it}) + \mu_i + \mu_t + u_{it} \quad (5)$$

As explanatory variables, the model uses the stock of past explorative words (EXP^*_{it}), the stock of R&D expenditures (RD^*_{it}), and the stock of patents (PAT^*_{it}). As additional regressors, we include two indicator variables, $D_1^{RD_{it}>0}$ and $D_2^{PAT_{it}>0}$, equal to 1 if firm i reports positive R&D stock and positive patent stock in year t. We estimate several versions of the above model. In some of the specifications, we control for year-fixed (μ_t) effects and industry-fixed effects (μ_i), which allow for variation in Q over time and across industries. In alternative specifications, we include firm-fixed effects in lieu of industry-fixed effects. To facilitate the interpretation of the estimates we again use the summary statistics reported in Table A.1 in Appendix A.

We find a positive and significant relation between the language-based exploration indicator and firm value. In columns (1) and (2), we control for industry- and year-fixed effects. The estimated coefficient of 0.200 on the exploration stock in model (1) shows that a one standard deviation increase in the stock of explorative words scaled by assets is linked to an increase in firm value by around 6.34 percent (or $e^{0.200\times0.3074}-1$). The relation remains statistically and economically significant even after we control for the level of R&D and patent stock of the firm. The estimated coefficient of 0.161 on the exploration stock variable from model (2) now shows that a one standard deviation increase in the stock of explorative words is associated with an increase in firm value by around 5.07 percent (or $e^{0.161\times0.3074}-1$).

Examining the control variables in column (2), we find that increases in R&D stock and patent stock also increase firm value, findings consistent with prior studies. Examining the economic significance of the estimates, we find that a one standard deviation increase in R&D stock results in an increase in value by 2.62 percent ($e^{0.090\times0.2870}-1$) while a one standard deviation increase in patent stock results in an increase in firm value by 7.74 percent ($e^{0.554\times0.1345}-1$). The empirical evidence, therefore, shows that exploration activities represented by explorative words are as important as R&D and patenting activities for firm value.

In columns (3) and (4), we repeat the analysis while accounting for firm- and year-fixed effects. We again find that the exploration stock variable has a positive and significant impact on firm valuation even after controlling for firm-fixed effects. The estimates in column (4) indicate that a one standard deviation increase in the stock of firm exploration increases a firm's Q by around 3.38 percent (or $e^{0.108\times0.3074}-1$).

To ensure the robustness of our findings, we also estimate an alternative non-linear specification, which we derive similarly to Hall et al. (2005). This specification is based on an additively separable linear specification (see also Griliches, 1981), where firm value is equal to:

$$V_{it} = q_t(A_{it} + \gamma K_{it})^{\sigma}. \quad (6)$$

In this specification, q_t is interpreted as the time-varying logarithmic average of Q, the γ parameter measures the shadow value of the stock of knowledge relative to the physical assets of the firm, and the σ parameter allows for non-constant scale effects. The two variables, A_{it} and K_{it}, measure the stock of physical capital and the stock of knowledge of firm i in year t. Taking the natural logarithm of both sides of Eq. (6), performing the necessary manipulations, and setting $Q_{it} = V_{it}/A_{it}$ we obtain:

$$\ln(Q_{it}) = \ln(q_t) + (\sigma-1)\ln(A_{it}) + \sigma\ln\left(1+\gamma\frac{K_{it}}{A_{it}}\right). \quad (7)$$

We assume that the stock of knowledge is a linear function of exploration stock, R&D stock, and patent stock. In addition, similar to Hall et al. (2005), we allow Q_{it} to differ systematically for firms with positive R&D stock and positive patent stock. Based on these assumptions, we estimate the following non-linear specification:

$$\ln(Q_{it}) = \ln(q_t) + (\sigma-1)\ln(A_{it}) + \sigma\ln\left(1+\gamma_1\frac{RD^*_{it}}{A_{it}}+\gamma_2\frac{PAT^*_{it}}{A_{it}}+\gamma_3\frac{EXP^*_{it}}{A_{it}}\right) + \delta_1 D_1^{RD^*_{it}>0} + \delta_2 D_2^{PAT^*_{it}>0} + u_{it} \quad (8)$$

Estimating Eq. (8) for the full sample of 88,175 observations using least squares, we obtain an adjusted R-squared of 51.68 percent. The estimates from this specification are shown in Eq. (9) below, with standard errors under each coefficient estimate, where standard errors are adjusted for firm-level clustering.

$$\ln(Q_{it}) = \ln(q_t) + \left(\underset{0.003}{1.000}-1\right)\ln(A_{it}) + \underset{0.003}{1.000}\ln\left(1+\underset{0.019}{0.237}\frac{RD^*_{it}}{A_{it}}+\underset{0.049}{0.505}\frac{PAT^*_{it}}{A_{it}}+\underset{0.014}{0.119}\frac{EXP^*_{it}}{A_{it}}\right) + \underset{0.012}{0.208}D_1^{RD^*_{it}>0} + \underset{0.011}{0.047}D_2^{PAT^*_{it}>0} + u_{it} \quad (9)$$

For the sake of brevity, we do not report the annual estimates of q_t, which vary between 0.990 in 2008 and 1.422 in 2017. As can be seen from the standard errors, all coefficients are significant at the 0.01 level. As a side note, the estimate of σ is equal to 1.00, implying constant returns to scale, a common assumption in the above-cited literature.

We highlight the positive and significant coefficient on exploration stock in relation to assets, which is again consistent with the idea that the stock market capitalizes exploration in the value of the firm. For an economic interpretation, we again examine a one standard deviation increase in the exploration stock indicator. Because the model is non-linear, the sensitivity of firm value to the exploration stock indicator depends on the levels of all knowledge-related variables. For the sake of this experiment, we assume all knowledge-related variables are at their mean levels (reported in Table A.1 in Appendix A). In such a case, a one standard deviation (or 0.6482) increase in the exploration stock variable from its mean value of 0.2525 results in an increase in firm value by around 6.71 percent.

The overall findings in this section show that firms that introduce more explorative words are valued higher by the market than firms that introduce fewer explorative words. Moreover, the relation between firm exploration and firm value is both statistically and economically significant. Combining these findings with the findings presented in the previous section, we conclude that, even though exploration increases firm risk, the benefits it creates through new growth opportunities outweigh the costs associated with the increased risk.

5. Conclusions

Exploration by firms has been recognized as a key factor shaping the accumulation of knowledge, early-stage innovation, and economic development. Motivated by a large literature emphasizing the importance of R&D in driving technological change, economists have traditionally used R&D spending to measure the exploratory efforts of firms. But economists have also acknowledged that the definition of industrial R&D may be too restrictive and that much can be learned by using a wider range of indicators related to the exploratory and innovative activities of firms (see among many others Pavitt, 1982, Freeman and Soete, 2009 and the references therein).

In this article, we propose a new language-based methodology for measuring creative exploration. Our approach fits within a broader set of recent endeavors using textual analysis to capture otherwise difficult to measure economic phenomena, such as product similarities, patent novelty, and management attitudes revealed during earnings calls. We contend that changes in natural language, specifically the creation of new words or the adoption of existing words in new settings, can be used to measure firm experimentation with new courses of action. To operationalize this idea, we create dictionaries of all nouns used by firms in their business descriptions and construct the innovations of these dictionaries over time. We define our proposed exploration indicator as the number of all new words adopted by a firm in a given year that were not widely used among the firm's peers by the time of adoption. Applying our approach to the business descriptions in mandatory Form 10-K annual filings, we create an exploration indicator at the firm-year level for all publicly traded firms in the U.S. between 1997 and 2017.

A range of analyses validate the proposed exploration indicator. For example, we show that our language-based exploration indicator predicts future R&D spending at the firm level. We also show that the exploration indicator is positively related to firm engagement with cutting-edge technologies. Based on the cogent arguments put forward by existing literature, exploration should lead to knowledge accumulation (Romer, 1994; Foster and Rosenzweig, 1995; Aghion and Howitt, 1998b). We confirm these expectations and show that our exploration indicator predicts patent and trademark applications as well as grants over the next three years. Also consistent with economic theory, we find that the exploration indicator is negatively related to firm age and positively related to firm risk.

The proposed exploration indicator contributes to a vast and growing literature on innovation indicators (for reviews, see Kleinknecht et al., 2002; Becheikh et al., 2006; Dziallas and Blind, 2019; Taques et al., 2021). Innovation indicators are typically analysed along multiple dimensions: the stage of the innovation process (early-stage or late-stage), the degree of novelty (exploration or exploitation), the degree of directness (direct or indirect), or the stage in the production process (input, intermediate, or output), among others. In an analysis of scientific publications between 1980 and 2015, Dziallas and Blind (2019) find that there are fewer indicators for the early stages of the innovation process than there are for the late stages and that early-stage innovation indicators are more commonly qualitative rather than quantitative. The paucity of formal indicators for early-stage innovation can be attributed to the difficulty of evaluating early ideas and concepts and to the dynamic and often unstructured nature of early-stage innovation. The exploration indicator proposed here contributes to the literature on quantifying early-stage innovation.

In addition, many useful existing indicators attempt to quantify firm innovation efforts at a granular level (e.g., new product announcements, amount of time spent on new ideas, number of failed ideas, Information and Communication Technology expenditures). However, as the Oslo Manual (OECD, 2018) observes, the administrative data required to construct many of the proposed innovation indicators are unavailable even for publicly listed firms because the data could be commercially sensitive and not subject to mandatory disclosures. Even indicators constructed using data from specially designed firm-level surveys, such as the Eurostat Community Innovation Survey, are available only for a limited number of industries and countries (OECD, 2015). In comparison, evaluation of innovation activities within our framework has several desirable properties. Our approach (i) can be applied to all publicly traded firms, (ii) is easy to replicate and extend, (iii) allows for a timely assessment of innovation efforts at the firm level, (iv) allows for comparability across firms, industries, or even geographic regions, and (v) allows for different levels of aggregation or disaggregation.

The intention of this study has not been to replace traditional indicators like R&D, but to provide an exploration indicator that captures different dimensions of the creative and innovation process. Our contributions in that aspect are several. Unlike R&D spending and other indicators of the creative and innovative processes, which are uncommon for many firms and many sectors of the economy, our exploration indicator covers all publicly listed firms in all industrial sectors. For example, while R&D spending is reported by less than 40 percent of all firm-years and tends to underestimate innovative efforts of services firms, our exploration indicator captures firm-level exploratory activities for all firms. In addition, the proposed indicator can capture a wide range of exploratory activities that are above and beyond those related to R&D activities. Following Schumpeter's (1934) framework, for instance, the explorative activity associated with each new word can be categorized into product, process, market, supply, and organization. Through a series of tests, we also demonstrate that our exploration indicator has different dynamic properties from traditional indicators. Specifically, we find that our exploration indicator tends to precede R&D spending and both patent and trademark applications as well as patent and trademark grants. Consistent with the idea that our indicator captures early-stage exploration efforts, we also find that R&D and patent and trademark grants do not predict future variations in the exploration indicator. We hope future research will investigate in even greater detail the timelines of specific new words and the innovations to which they refer, including corresponding R&D spending, patenting, and trademarking activities.

Our approach is based on the assumption that corporate disclosures adequately reflect firm activities of material importance. But one may reasonably question this assumption and ask whether existing frictions and agency issues prompt firms to report their exploratory activities in a distorted manner. On the one hand, companies could adopt new rhetorical words or symbolisms to make themselves look better, leading to an over-reporting of exploratory activities. On the other hand, companies could under-report exploration due to a need for conformity or secrecy. For example, it is well understood that reputational concerns create incentives for agents to conform their behavior to the behavior of others. This is famously summarized by Keynes (1936), who noted that "(w)orldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally." Exploration is indeed new, risky, and unconventional. We have discussed several reasons, including existing laws and regulations, justifying why our proposed indicator contains material information related to firm exploration.

In our empirical analysis, we have further demonstrated the usefulness of our framework by examining a natural and important question: Does exploration add economic value? To answer this question, we construct an indicator for the cumulative stock of exploration and find that this indicator is positively related to firm value even after controlling for traditional variables such as the stock of past R&D spending and patenting activities. Moreover, we find that the proposed indicator's relation to firm value is economically large so that a one standard deviation increase in the exploration stock indicator is associated with a roughly five percent increase in firm value.

In the end, we would like to note that our approach can be used to examine other questions of interest to economists, such as the role of intellectual spillovers or industrial agglomeration in the innovative process, for example. More generally, our language-based approach can be applied to measure creative contributions in other domains, such as government grant applications, academic publications, and even literary works of art.

CRediT authorship contribution statement

Vladimir A. Gatchev: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Project administration. **Christo A. Pirinsky:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Buvaneshwaran Venugopal:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful for the valuable comments and suggestions made by the Editor, Ben R. Martin, and three anonymous reviewers for this journal. For helpful suggestions we would like to thank Vladimir Atanasov, Dan Bradley, Honghui Chen, Val Dimitrov, Ioannis Floros, Gerard Hoberg, Candace Jens, David McLean, Neale Mahoney, Manpreet Singh, Qinghai Wang, and participants at webinars hosted by University of Central Florida and Tulane University.

Appendix

Tables A1–A3.

Table A2
Examples of explorative words for two sectors.

#	IT Software & Services	Materials	#	IT Software & Services	Materials
1	software-as-a-service	footprint	26	tokenization	attestation
2	gdpr	dowdupont	27	Byod	biodiesel
3	android	sustainability	28	openstack	chemtura
4	ipad	ethic	29	ransomware	intimacy
5	iphone	irma	30	alcatel-lucent	pre-recession
6	twitter	legacy	31	google	axiall
7	facebook	stimulus	32	on-boarding	headwind
8	blog	biofuel	33	platform-as-a-service	end-market
9	devops	endangerment	34	consumerization	dismissal
10	html5	euro	35	anti-corruption	3d
11	omnichannel	macro	36	vantiv	patriot
12	cfpb	emea	37	eprivacy	otcqb
13	blockchain	agility	38	Iiot	segmentation
14	portal	venator	39	ecosystem	mexichem
15	cloud	homeland	40	curate	chemours
16	linkedin	easement	41	instagram	compromise
17	hadoop	arcelormittal	42	gamification	ineo
18	millennial	holdco	43	governance	rescue
19	vmware	ingevity	44	cybersecurity	differentiator
20	onboarding	cyberattack	45	infor	amex
21	codm	lomon	46	hyper-v	codm
22	centurylink	Lord	47	netbook	jiangsu
23	fireeye	trajectory	48	leidos	lanxess
24	virtualization	debtor-in-possession	49	cryptocurrency	bact
25	infrastructure-as-a-service	codification	50	csra	fitesa

Notes: The table provides examples of explorative words that were introduced and subsequently adopted in the IT Software & Services sector and the Materials sector. We rank these words in terms of subsequent adoption frequency, with highly adopted words listed first.

Table A1
Summary statistics of the main variables.

	Observations	Mean	Standard deviation
Exploration indicator	88,175	7.4010	8.4723
Exploration stock / Assets	88,175	0.2525	0.6482
ln(1 + Exploration stock / Assets)	88,175	0.1595	0.3074
Citation-weighted patent applications	88,175	3.7784	16.8442
ln(1 + Citation-weighted patent applications)	88,175	0.3908	0.9917
Citation-weighted patent grants	88,175	5.0524	21.1040
ln(1 + Citation-weighted patent grants)	88,175	0.4847	1.0997
Patent applications stock / Assets	88,175	0.0564	0.1981
ln(1 + Patent applications stock / Assets)	88,175	0.0436	0.1345
Trademark applications	88,175	2.1056	5.0050
ln(1 + Trademark applications)	88,175	0.5719	0.8843
Trademark grants	88,175	1.3593	3.1118
ln(1 + Trademark grants)	88,175	0.4628	0.7476
R&D expenditures (mill. 2018 USD)	88,175	31.1002	118.9282
ln(1 + R&D expenditures)	88,175	1.1819	1.7812
R&D stock / Assets	88,175	0.2231	0.5260
ln(1 + R&D stock / Assets)	88,175	0.1483	0.2870
Assets (mill. 2018 USD)	88,175	4,201.3381	11,700.0000
ln(Assets)	88,175	6.3608	2.0874
Beta	88,144	1.0778	0.9172
Idiosyncratic volatility (%)	88,075	6.5938	4.2649
Market-to-book	88,175	1.8511	1.4398
ln(Market-to-book)	88,175	0.4297	0.5544

Table A3
Dynamics between the exploration indicator and R&D spending for manufacturing and service firms.

Dependent variable: ln(1 + R&D $_t$)	Full sample (1)	Manufacturing firms (2)	Service firms (3)
Exploration indicator $_t$	0.036	0.038	0.045
	(0.030)	(0.053)	(0.032)
Exploration indicator $_{t-1}$	0.122***	0.170***	0.097***
	(0.029)	(0.051)	(0.030)
Exploration indicator $_{t-2}$	0.089***	0.136***	0.082***
	(0.027)	(0.049)	(0.029)
Exploration indicator $_{t-3}$	0.057**	0.146***	− 0.005
	(0.026)	(0.047)	(0.027)
ln(Assets $_t$)	24.489***	45.919***	10.829***
	(0.386)	(0.697)	(0.430)
Firm- & year-fixed effects	Yes	Yes	Yes
Observations	53,135	20,272	30,207
Adjusted R-squared (%)	96.20	95.24	95.33

Notes: Manufacturing firms are those firms within division D (Manufacturing) of the Standard Industrial Classification system. Service firms are those firms within divisions E (Transportation, Communications, Electric, Gas, and Sanitary Services), F (Wholesale Trade), G (Retail Trade), H (Finance, Insurance, and Real Estate), I (Services), and J (Public Administration). Standard errors in parentheses *** $p<0.01$, ** $p<0.05$, *** $p<0.10$.

References

Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. Econometrica 60, 323–351.
Aghion, P., Howitt, P., 1998a. Capital accumulation and innovation as complementary factors in long-run growth. J. Econ. Growth 3, 111–130.
Aghion, P., Howitt, P., 1998b. Endogenous Growth Theory. MIT Press, Cambridge, MA.
Akcigit, U., Kerr, W.R., 2018. Growth through heterogeneous innovations. J. Polit. Econ. 126, 1374–1443.
Algeo, J., 1993. Fifty Years Among the New Words: A Dictionary of Neologisms, 1941-1991. Cambridge University Press.
Ang, J.S., Cole, R.A., Lin, H.W., 2000. Agency costs and ownership structure. J. Finance 55, 81–106.
Arrow, K.J., 1962. The economic implications of learning by doing. Rev. Econom. Stud. 29, 155–173.
Barge-Gil, A., López, A., 2014a. R&D determinants: accounting for the differences between research and development. Res. Policy 43, 1634–1648.
Barge-Gil, A., López, A., 2014b. R versus D: estimating the differentiated effect of research and development on innovation results. Industrial and Corporate Change 24, 93–129.
Becheikh, N., Landry, R., Amara, N., 2006. Lessons from innovation empirical studies in the manufacturing sector: a systematic review of the literature from 1993 to 2003. Technovation 26, 644–664.
Belenzon, S., Patacconi, A., 2013. Innovation and firm value: an investigation of the changing role of patents, 1985-2007. Res. Policy 42, 1496–1510.
Benner, M.J., Tushman, M.L., 2003. Exploitation, exploration, and process management: the productivity dilemma revisited. Acad. Manage. Rev. 28, 238–256.
Bernstein, S., 2015. Does going public affect innovation? J. Finance 70, 1365–1403.
Blankespoor, E., Hendricks, B.E., Miller, G.S., 2017. Perceptions and price: evidence from CEO presentations at IPO roadshows. J. Accounting Res. 55, 275–327.
Bloom, N., Van Reenen, J., 2002. Patents, real options and firm performance. Econom. J. 112 (478), C97–C116.
Brown, J.R., Fazzari, S.M., Petersen, B.C., 2009. Financing innovation and growth: cash flow, external equity, and the 1990s R&D boom. J. Finance 64, 151–185.
Carpenter, R.E., Petersen, B.C., 2002. Capital market imperfections, high-tech investment, and new equity financing. Econom. J. 112 (477), 54–72.
Chen, X., Han, T., 2019. Disruptive technology forecasting based on Gartner hype cycle. In: 2019 IEEE Technology & Engineering Management Conference. IEEE, pp. 1–6.
Crystal, D., 2002. The English Language: A Guided Tour of the Language. Penguin Books UK.
Czarnitzki, D., Hottenrott, H., Thorwarth, S., 2011. Industrial research versus development investment: the implications of financial constraints. Cambridge J. Econ. 35, 527–544.
Czarnitzki, D., Hottenrott, H., Thorwarth, S., 2009. The knowledge production of 'R' and 'D. Econ. Lett. 105, 141–143.
Dedehayir, O., Steinert, M., 2016. The hype cycle model: a review and future directions. Technol. Forecast Soc. Change 108, 28–41.
Dimitrov, V., Tice, S., 2006. Corporate diversification and credit constraints: real effects across the business cycle. Rev. Financ. Stud. 19, 1465–1498.
Dinlersoz, E., Goldschlag, N., Myers, A., Zolas, N., 2019. An natomy of U.S. firms seeking trademark registration. In: Corrado, C., Haskel, J., Miranda, J., Sichel, D. (Eds.), Measuring and Accounting for Innovation in the 21st Century. University of Chicago Press.
Drejer, I., 2004. Identifying innovation in surveys of services: a Schumpeterian perspective. Res. Policy 33, 551–562.
Dukakis, A., 2020. Quest Diagnostics has Capacity to Run 25,000 Coronavirus Tests a Day. CEO. ABC News Coronavirus Health & Science. March 26, 2020.
Dyke, C.V., 1992. Old words for new worlds: modern scientific and technological word-formation. Am. Speech 67, 383–405.
Dziallas, M., Blind, K., 2019. Innovation indicators throughout the innovation process: an extensive literature analysis. Technovation 80-81, 3–29.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. J. Finance 47, 427–465.
Faurel, L., Li, Q., Shanthikumar, D.M., Teoh, S.H., 2021. Product development innovation: insights from trademarks. SSRN Working Paper No. 2585804.
Foster, A.D., Rosenzweig, M., 1995. Learning by doing and learning from others: human capital and technical change in agriculture. J. Polit. Econ. 103, 1176–1209.
Francis, J., Smith, A., 1995. Agency costs and innovation: some empirical evidence. J. Accounting Econom. 19, 383–409.
Freeman, C., Soete, L., 2009. Developing science, technology and innovation indicators: what we can learn from the past. Res. Policy 38, 583–589.
Geroski, P.A., Van Reenen, J., Walters, C.F., 1997. How persistently do firms innovate? Res. Policy 26, 33–48.
Griliches, Z., 1981. Market value, R&D, and patents. Econ. Lett. 7, 183–187.
Griliches, Z., 1986. Productivity, R&D, and basic research at the firm level in the 1970's. Am. Econ. Rev. 76, 141–154.
Griliches, Z., 1990. Patent statistics as economic indicators: a survey. J. Econ. Lit. 28, 1661–1707.
Griliches, Z., Pakes, A., Hall, B.H., 1986. The value of patents as indicators of inventive activity. NBER Working Paper No. 2083.
Grossman, G.M., Helpman, E., 1991. Innovation and Growth in the Global Economy. MIT Press.
Grossman, G.M., Shapiro, C., 1982. A theory of factor mobility. J. Polit. Econ. 90, 1054–1069.
Gu, L., 2016. Product market competition, R&D investment, and stock returns. J. financ. econ. 119, 441–455.
Hall, B., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. RAND J. Econ. 36, 16–38.
Hannan, M.T., Freeman, J., 1984. Structural inertia and organizational change. Am. Sociol. Rev. 49, 149–164.
Hassan, T.A., Hollander, S., van Lent, L., Tahoun, A., 2019. Firm-level political risk: measurement and effects. Q. J. Econ. 134, 2135–2202.
Heading, R.C., 2017. Proton pump inhibitor failure in gastro-oesophageal reflux disease: a perspective aided by the Gartner hype cycle. Clin. Med. 17, 132–136.
Helwege, J., Pirinsky, C., Stulz, R.M., 2007. Why do firms become widely held? An analysis of the dynamics of corporate ownership. J. Finance 62, 995–1028.
Henderson, V.J., Storeygard, A., Weil, D.N., 2012. Measuring economic growth from outer space. Am. Econ. Rev. 102, 994–1028.
Hoberg, G., Phillips, G., 2016. Text-based network industries and endogenous product differentiation. J. Polit. Econ. 124, 1423–1465.
Hobijn, B., Jovanovic, B., 2001. The information-technology revolution and the stock market: evidence. Am. Econ. Rev. 91, 1203–1220.
Holmström, B., 1989. Agency costs and innovation. J. Econ. Behav. Organ. 12, 305–327.
Irwin, D.A., Klenow, P.J., 1994. Learning-by-doing spillovers in the semiconductor industry. J. Polit. Econ. 102, 1200–1227.
Jaffe, A.B., Trajtenberg, M., 2002. Patents, Citations, and Innovations: A Window On the Knowledge Economy. MIT press.
Jaravel, X., Petkova, N., Bell, A., 2018. Team-specific capital and innovation. Am. Econ. Rev. 108, 1034–1073.
Jarvenpaa, H.M., Makinen, S.J., 2008. Empirically detecting the hype cycle with the life cycle indicators: an exploratory analysis of three technologies. In: 2008 IEEE International Conference on Industrial Engineering and Engineering Management. IEEE, pp. 12–16.
Jovanovic, B., Nyarko, Y., 1996. Learning by doing and the choice of technology. Econometrica 64, 1299–1310.
Kelly, B., Papanikolaou, D., Seru, A., Taddy, M., 2021. Measuring technological innovation over the long run. Am. Econom. Rev.: Insights 3, 303–320.
Kerr, W.R., Nanda, R., Rhodes-Kropf, M., 2014. Entrepreneurship as experimentation. J. Econ. Perspect. 28, 25–48.
Keynes, J.M., 1936. The General Theory of Employment, Interest and Money. Harcourt, Brace and Company, New York, NY.
Kleinknecht, A., Van Montfort, K., Brouwer, E., 2002. The non-trivial choice between innovation indicators. Econom. Innov. New Technol. 11, 109–121.
Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Q. J. Econ. 132, 665–712.
Lerner, J., Seru, A., 2021. the use and misuse of patent data: issues for corporate finance and beyond. Rev. Financ. Stud., Forthcoming. https://doi.org/10.1093/rfs/hhab084.
Loughran, T., McDonald, B., 2014. Measuring readability of financial disclosures. J. Finance 69, 1643–1671.
Jr, Lucas, E, R., Moll, B., 2014. Knowledge growth and the allocation of time. J. Polit. Econ. 122, 1–51.
Manso, G., 2011. Motivating innovation. J. Finance 66, 1823–1860.
March, J.G., 1991. Exploration and exploitation in organizational learning. Organization Sci. 2, 71–87.
Martimort, D., 1999. The life cycle of regulatory agencies: dynamic capture and transaction costs. Rev. Econom. Stud. 66, 929–947.
McDonald, L., 2005. The meaning of e-: neologisms as markers of culture and technology. TOPIA: Canad. J. Cultural Stud. 14, 82–90.
Mendonça, S., Pereira, T.S., Godinho, M.M., 2004. Trademarks as an indicator of innovation and industrial change. Res. Policy 33, 1385–1404.
Nelson, R.R., Winter, S.G., 1982. An Evolutionary Theory of Economic Change. Belknap Press of Harvard University Press.
O'Leary, D.E., 2008. Gartner's hype cycle and information system research issues. Int. J. Accounting Information Syst. 9, 240–252.
OECD, 2015. The measurement of scientific, technological and innovation activities. Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development. OECD Publishing, Paris (available at. http://doi.org/10.1787/9789264239012-en).
OECD, 2018. The measurement of scientific, technological and innovation activities. Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data On Innovation, 4th Edition. OECD Publishing, Paris/Eurostat, Luxembourg. https://doi.org/10.1787/9789264304604-en). available at:
Pakes, A., 1985. On patents, R&D, and the stock market rate of return. J. Polit. Econ. 93, 390–409.
Pastor, L., Veronesi, P., 2009. Technological revolutions and stock prices. Am. Econ. Rev. 99, 1451–1483.
Pavitt, K., 1982. R&D, patenting and innovative activities. Res. Policy 11, 33–51.
Pavitt, K., 1984. Sectoral patterns of technical change: towards a taxonomy and a theory. Res. Policy 13, 343–373.

Pavitt, K., Robson, M., Townsend, J., 1989. Technological accumulation, diversification, and organisation in UK companies, 1945-1983. Manage. Sci. 35, 81–99.

Princeton University, 2010. About WordNet, WordNet. Princeton University.

Quintana-García, C., Benavides-Velasco, C., 2008. Innovative competence, exploration and exploitation: the influence of technological diversification. Res. Policy 37, 492–507.

Rogers, E.M., 1962. Diffusion of Innovations. Free Press of Glencoe.

Romer, P.M., 1986. Increasing returns and long-run growth. J. Polit. Econ. 94, 1002–1037.

Romer, P.M., 1994. The origins of endogenous growth. J. Econ. Perspect. 8, 3–22.

Schumpeter, J.A., 1934. The Theory of Economic Development. Harvard University Press, Cambridge, MA.

Simeth, M., Cincera, M., 2016. Corporate science, innovation, and firm value. Manage. Sci. 62, 1970–1981.

Steinert, M., Leifer, L.J., 2010. Scrutinizing Gartner's hype Cycle approach, PICMET 2010 Technology Management For Global Economic Growth. IEEE, pp. 1–13.

Taques, F.H., López, M.G., Basso, Leonardo F., Areal, Nelson, 2021. Indicators used to measure service innovation and manufacturing innovation. J. Innov. Knowl. 6, 11–26.

Young, H.P., 2009. Innovation diffusion in heterogeneous populations: contagion, social influence, and social learning. Am. Econ. Rev. 99, 1899–1924.