s:fi
RESEARCH

Swiss Finance Institute Research Paper Series N°19-22

Rapidly Evolving Technologies and Startup Exits



Donald Bowen:
Lehigh University

Laurent Frésard:
Università della Svizzera italiana and Swiss Finance Institute

Gerard Hoberg:
University of Southern California

Rapidly Evolving Technologies and Startup Exits

Donald Bowen, Laurent Frésard, and Gerard Hoberg.*

November 24, 2021

ABSTRACT

This paper examines startups positioning within technological cycles. We use patent text to measure whether innovation pertains to a technological area that is rapidly evolving or stable. We show that innovation in rapidly evolving areas (i.e., early in the cycle) substitute for existing technologies, whereas innovation in stable areas (i.e., later in the cycle) complement them. Our new measure is distinct from existing characterizations of innovation and is economically important. We find that startups in rapidly evolving areas tend to exit via IPO, thus remaining independent, consistent with technological substitution. In contrast, startups in stable areas tend to sell-out, consistent with technological complementarity and synergies.

Key words: Patent Text, Technology Waves, Innovation, Technology Substitution, Venture Capital, Startup Exit, Initial Public Offerings (IPOs), Acquisitions, Sell-Outs

JEL classification: G32, G34, G24

*Lehigh University, the Universita della Svizzera Italiana (Lugano), Swiss Finance Institute, and the University of Southern California respectively. Bowen can be reached at deb219@lehigh.edu, Frésard can be reached at laurent.fresard@usi.ch, and Hoberg can be reached at hoberg@marshall.usc.edu. We thank Sunil Muralidhara and Lauren Phillips for excellent research assistance. We also thank Jay Ritter on sales data for IPO firms. For helpful comments, we thank Nick Bloom, Francesco D'Acunto, Francois Degeorge, Francois Derrien, Michael Ewans, Joan Farre-Mensa, Francesco Franzoni, Jack He, Sabrina Howell, Mark Lang, Josh Lerner, Michelle Lowry, Song Ma, Adrien Matray, Dimitris Papanikolaou, Shawn Thomas, Xuan Tian and participants at American University, Arizona State Univesity, Babson, Bocconi University, Boston College, Dauphine University, the CCF conference in Bergen, the CETAFE conference at the University of Southern California, the FinTech Conference at Georgia State University, the FRA conference in Las Vegas, Lehigh University, The Midwest Finance Association meeting in Chicago, the NBER meeting on Productivity, Innovation, and Entrepreneurship, Georgetown University, HEC Paris, Johns Hopkins University, San Diego University, the SFI Research days, Stockholm School of Economics, Universita Cattolica, University of Virginia, Texas Christian University, Toulouse School of Economics, Tsinghua University, CKGSB, The UBC Winter Conference in Whistler, the Universita della Svizzera Italiana (Lugano), the University of Maryland, the University of Zurich, the Western Finance Association meeting in Huntington Beach, and Virginia Tech. This paper builds on parts of a previous version titled "Technological Disruptive Potential and the Evolution of IPO and Sell-outs". All errors are the authors' alone. All rights reserved by Donald Bowen, Laurent Frésard, and Gerard Hoberg.

I Introduction

Technological areas tend to follow cycles, initiated by a technological breakthrough that creates a discontinuity with existing practices (Tushman and Anderson (1986)). The breakthrough is typically followed by a fluid phase during which the technology evolves rapidly (and in various directions) as a result of widespread experimentation and trial-and-error (Callander (2011)), eventually reaching a stable state until a new breakthrough occurs (Abernathy and Utterback (1978)). These cycles, which differ in frequency and length across technology areas, govern the speed and direction of technological changes and determine the nature of innovative activities. Indeed, innovation occurring during a breakthrough period typically aims at substituting existing technologies (Mokyr (1990)), whereas innovation developed in the stable phase of cycles tends to be more incremental and complementary (Klepper (1996)). Thus, the positioning of firms in technology cycles should affect the characteristics of their innovation, the associated sources of economic value, and its organization.

In this paper, we develop a new methodology to measure the positioning of firms in technology cycles. We use the text of firms' patents to identify whether their patented innovation pertains to a technological area that is rapidly evolving (i.e., following breakthroughs) or stable. We show that this text-based measure provides a novel characterization of the technological innovation cycle that is distinct from existing measures. We then demonstrate the economic relevance of our measure by studying how the distinction between innovation in rapidly evolving or stable technology areas relates to startups' modes of exit (i.e., going public or selling out to another company).

We track the evolution of technology areas using the text from all patents filed with the U.S. Patent and Trademark Office (USPTO) between 1930 and 2010 (6.6 million patents). To capture the positioning of a given patent within technology cycles, we measure the intensity with which its vocabulary is growing in usage across all recent and contemporary patents. We label this continuous variable "RETech" (short for "Rapidly Evolving Technology") and define patents as being in technology areas that are evolving rapidly if they rely on vocabulary that is growing rapidly in the patent corpus overall. Conversely,

we define patents as belonging to stable technology areas if they use vocabulary that is not growing rapidly.

To illustrate the logic of our measure, consider typical genetics words such as "peptide", "clone", "nucleic", or "recombinant". Their usage in patents soared in 1995, reflecting rapid breakthroughs in genome sequencing technology. Our measure classifies patents using such words in 1995 as pertaining to a rapidly evolving technology area, reflecting the underlying growth in the genomics technology they contribute to. Yet, we classify patents using the exact same words in 2005 as being in a stable technology area, as genomic technologies had matured and the related vocabulary stabilized.

Remarkably, RETech is only weakly correlated with a host of existing innovation measures (e.g., citations, originality, or scope), indicating that it captures independent information about the characteristics of technological innovation. In addition, we show that patents recognized by the USPTO as historical breakthroughs (e.g., television, computer, helicopter, and advances in modern genetics) occurred in rapidly evolving areas (i.e., high values of RETech) and that a patent's RETech is positively related to the stock market reaction to its approval by the USPTO. Furthermore, unlike other existing measures of innovation (e.g., those based on future citations), RETech is *ex ante* measurable as it only relies on prior information to identify patents focusing on rapidly growing vocabularies.[1]

RETech builds on the idea that patents in rapidly evolving areas aim to substitute existing technologies. A direct implication is that these patents should "crowd out" the prior art they cite in terms of future cites. Following Funk and Owen-Smith (2016), we capture such crowding-out by examining the extent to which future patents cite a focal patent but do not cite any of the patents cited by the focal patent. Validating the economic logic of our measure, we show that the likelihood of technological crowding-out varies strongly and positively with RETech, as it is 20% higher for patents in the top decile than in the middle decile, which is in turn 5% higher than for patents in the lowest decile. We find the reverse as well: Patents in stable areas (low values of RETech) are much more likely to complement their predecessors, as indicated by cases where future

[1]The absence of look-ahead bias is important for applications seeking to explain firms' or inventors' decisions based on timely patent information, thereby reducing endogeneity and truncation concerns.

patents cite a focal patent *and* at least one of the patents it cited.

To further establish the economic relevance of RETech, we study its associations with startups' exits. A successful exit is one of the most important events in the life of a startup, and success is typically achieved when a startup sells shares to outside investors during an initial public offering (IPO) or when it directly sells out to another company. We posit that startups innovating in rapidly evolving technology areas have high stand-alone values because their technologies tend to substitute existing technologies. These startups can create an independent market presence with substantial growth potential by substituting and replacing established technologies, without partnering with an acquirer and sharing rents. Hence, these startups should favor exiting via an independent public listing. In contrast, startups innovating in stable technology areas likely offer significant synergies to potential buyers, because as noted above, innovation in stable areas tends to complement existing technology. These startups should favor merging with another entity as the gains from synergies should more than offset the cost of sharing future payoffs with a buyer.

To verify if startups' exits are indeed related to their positioning in technology cycles, we consider a large sample of 9,167 U.S. venture-backed startups from 1980-2010 and their 94,703 patents, which we aggregate to obtain a measure of RETech at the startup-quarter level.[2] We show that startups' RETech is strongly related to their exit choices. In univariate tests, we find that the RETech of an average IPO startup is 40% higher than that of startups exiting via a sell-out. In multivariate tests, we estimate that startups innovating in rapidly evolving technological areas are significantly more likely to go public and are less likely to exit by selling out. A one standard deviation increase in a startup's ex ante RETech is associated with a 24.7% larger probability of exit via IPO in the next quarter (i.e., from a baseline probability of 0.42% to 0.52%) and a 18.8% lower likelihood of a sell-out (from 0.73% to 0.62%).

Because RETech could arguably be related to other (observed and unobserved) deter-

[2]VC-backed startups are a relevant sample for two reasons. First, data is available to link exit timing to patent data. Second, VC-backed startups are a large fraction of the IPO and acquisition market (Ritter (2019)) and produce a large share of total innovation (Gornall and Strebulaev (2015)).

minants of startup exit (e.g., their quality), we show that the role of RETech remains robust when we control for a host of startup and market characteristics, such as startup age, the size of their patent portfolio, whether they have raised substantial external finance, the characteristics of the lead VC, and overall market conditions. Our results are also robust to controlling for other technology traits such as technological "breadth" (patents using vocabulary from diverse bodies of knowledge), patents' similarity to other firms, citations, originality, number of claims, and scope. The inclusion of stringent fixed effects based on the startup's vintage, age, geographical location, industry, and technological categories also does not alter our findings. Overall, we find strong and robust associations that confirm our predictions regarding technological life cycles and exit methods. However, our results do not establish causality, indicating a fruitful direction for future research.

To further explore economic mechanisms, we test if the role of RETech in startup exits is stronger when the economic benefits of technology substitution are higher. We propose that these benefits should be larger in established technology areas, because startups innovating in rapidly evolving established technologies (i.e., existing areas experiencing new waves of innovation) can likely capitalize on potentially large existing markets. Dividing patent vocabularies into new and established technology areas based on the age of each word (relative to when it first appears in the patent corpus), we confirm that the role of rapidly evolving technologies is more important in predicting a higher incidence of IPOs and a lower incidence of sell-outs in established areas than in nascent areas.

In addition, we provide three sets of ancillary findings that further support the importance of substitution and complementarity. First, we predict that combining complementary technologies is likely most valuable when startups and acquirers are both innovating in stable technology areas. In a sub-sample of startups selling out to public firms owning patents (i.e. buyers for whom we can measure RETech), we show that the majority of these sell-outs involve startups and acquirers both innovating in stable technology areas. Second, we show that investor announcement returns to sell-outs, a widely-used measure of expected synergies, are significantly more positive when the targeted startups innovate

in stable areas. Third, focusing on startups exiting via IPO, we find that those innovating in more rapidly evolving areas exit faster and raise additional new capital in lieu of selling insider equity (i.e., issue more primary shares) during their IPO. This is consistent with needing larger amounts of funding in order to quickly capitalize on substituting technologies (and build the necessary organizational capital to facilitate independent market positions).[3] Overall, our results are consistent with a role for substitution and complementarity, but we also note that exits are influenced by many economic theories (which we discuss later in Section IV.A), and our results are not mutually exclusive to them.

The results in this paper contribute to research developing patent-based measures to characterize technological innovation (Hall, Jaffe, and Trajtenberg (2001), Lerner (1994) or Kogan, Papanikolaou, Seru, and Stoffman (2016)), and in particular to recent methodological studies using patent text. Kelly, Papanikolaou, Seru, and Taddy (2019) measure patent "significance" based on textual similarity to prior and future patents. Packalen and Bhattacharya (2018) and Balsmeier, Assaf, Chesebro, Fierro, Johnson, Johnson, Li, Luck, O'Reagan, Yeh, Zang, and Fleming (2018) identify new ideas using the first appearance of words in patents. Bena, Ortiz-Molina, and Simintzi (2020) classify patents into process and non-process innovation. Our study develops a new measure identifying the positioning of patents in technology cycles. While we illustrate the usefulness of this new measure by focusing on startup exits, we note that its ability to identify technology waves and to distinguish patented innovation that substitutes or complements existing technologies may also be relevant to examine a range of other questions.[4]

Our results also add to the growing research studying the role played by patents in startups' development and success. Existing research reports that patenting in startups is associated with more funding, stronger growth, and more successful exits (Cockburn and MacGarvie (2009), Gaule (2018), or Farre-Mensa, Hegde, and Ljungqvist (2019)).[5] Whereas existing studies analyze the effects of the number of patents (or the first patent),

[3]We also find that, prior to exit, startups innovating in rapidly evolving technology areas receive more funding from VCs both in terms of the number of rounds and dollar amounts.

[4]For this reason, the authors will distribute RETech data and code publicly on their websites.

[5]Relatedly, other studies consider startups' patenting as a milestone measuring success (Bernstein, Giroud, and Townsend (2016)).

we focus on the technological specificities of startups' patents and study how their positioning in technology cycles determine firm boundaries and development.

Finally, our analysis also adds to the literature studying the determinants of startups' exit (see Da Rin, Hellman, and Puri (2013) for a survey). While most studies either examine IPOs and sell-outs in isolation, or bundle both into a single measure of "successful" exit (Bernstein, Giroud, and Townsend (2016) and Guzman and Stern (2015)), we examine both jointly and document a strong role for startups' technological characteristics. Our results add to studies indicating that exit choices are related to founders' private benefits of control, product market presence, and growth potential (Cumming and Macintosh (2003), Bayar and Chemmanur (2011), Poulsen and Stegemoller (2008), Chemmanur, He, He, and Nandy (2018), Ewens and Farre-Mensa (2020) or Chemmanur, He, Ren, and Shu (2020)).

II Data and New Text-Based Measures

A Patent Data and Text

We gather information from Google Patents for all 6,595,226 U.S. patents that were applied for between 1930 and 2010 and granted by 2013 by the USPTO. We gather the publication date, application date, name of inventor(s), and initial assignee(s). We also collect the full patent text and information on the technology classification of the patents by converting the U.S. Patent Classification (USPC) into the two-digit NBER technology codes created in Hall, Jaffe, and Trajtenberg (2001).[6] Since we are mostly interested in measuring the technological changes pertaining to the corporate sector, we categorize each patent into four types of applicants: U.S. public firms, U.S private firms, foreign (private or public) firms, or others (e.g., individuals, universities, or foundations). For brevity, we describe this classification method in the Internet Appendix (Section IA.A).

[Insert Figure I about here]

[6]The patent text web-crawled from Google is the text of the final granted patent and this sample only includes granted patents. The text of initial patent applications, including those not granted, is only available after 2001, which is why we focus on granted patents.

The full text of each patent consists of three distinct sections: abstract, claims, and description. The claims section defines the scope of legal protection granted. The description section explicitly describes the characteristics of the invention/innovation. The abstract contains a summary of the disclosure contained in the description and claims sections. Figure I presents an example of a typical patent's textual structure (#6285999, "A method for node ranking in a linked database", assigned to Google in 1998).

To capture the genuine technological characteristics of patented inventions from text, we only use the description section.[7] Indeed, practitioners, judges, and academics have noted that the claims section is materially impacted by lawyers and contains stilted legalistic language (Bena and Simintzi (2019)).[8] In contrast, the description section is required by law to explain the construction and use of the invention in "*full, clear, concise, and exact terms*" so that persons skilled in the area could recreate and use the invention (see 35 U.S. Code 112). Hence, although lawyers contribute to the description section, their impact is likely lower than for other sections.[9]

Following earlier studies constructing variables from text (Hanley and Hoberg (2010) or Hoberg and Phillips (2016)), we represent the text of each patent as a numerical vector. Let N_t denote the number of distinct words used in the description sections of all patent applications in a given year t, after removing common words appearing in more than 25% of all patents in a given year (this step also removes stop words). We organize patents based on their application year rather than their grant year, as this more accurately reflects the timing of innovation. Each patent j applied for in year t is represented by an ordered vector $V_{j,t}$ of length N_t, and each element corresponds to the number of times the corresponding word is used by patent j. If patent j does not use a given word, the

[7] Our results are robust although a bit weaker if we use the claims section instead of the description section.

[8] For example, Giles Rich, a judge on the Court of Appeals for the Federal Circuit and co-author of the 1952 Patent Act, famously wrote in 1990 that "The name of the game is the claim." Bena and Simintzi (2019) state "Claims are written using stilted legalistic language and consistent vocabulary across firms and over time" because of patent law (35 U.S. Code 112). Moreover, evaluating claims is the statutorily defined focus of examiners (37 CFR 1.104).

[9] As further evidence that our results are likely not driven by legal content, we note our results are also robust to controlling for lawyer fixed effects.

corresponding element is zero.[10]

Due to the large number of words used across all patents, these $V_{j,t}$ vectors are sparse. For instance, in 1980, the number of distinct words (after applying our filters) used in an average patent is 352 (median is 300), while there are N_{1980}=400,097 distinct words across all patents. Patents have gotten longer over time. As of 2000, the average and median are 453 and 338, and N_{2000} is 1,358,694.

B Identifying Patents in Rapidly Evolving Technology Areas

To measure the relative positioning of patents in technology cycles, we consider whether a patent pertains to technology areas that are rapidly evolving or stable. We do so based on whether the vocabulary used to describe the invention in a given patent is growing in usage across all existing patents.

Our measure is constructed in three steps.[11] First, we compute the aggregate vector z_t (of dimension N_t) which summarizes the use of each word in year t across all patents:

$$z_t = \frac{1}{|P_t|}\left(\sum_{j \in P_t} \frac{V_{j,t}}{V_{j,t} \cdot \mathbf{1}}\right), \tag{1}$$

where the operator "$\cdot$" denotes the dot product, $\mathbf{1}$ is a vector of ones, $\frac{V_{j,t}}{V_{j,t} \cdot \mathbf{1}}$ are vectors representing the normalized word distribution used by patent j (this normalization controls for the document length and ensures each patent has equal weight), P_t represents the set of all patents in year t, and $|P_t|$ is the number of patents in year t.

Second, we compute the relative difference in aggregate word usage between $t-1$ and t as

$$\Delta_t = \frac{z_t - z_{t-1}}{z_t + z_{t-1}}, \tag{2}$$

where the dimension of Δ_t is defined by the the union of words in year t and $t-1$ (i.e., N_t and N_{t-1}).[12] By construction, elements of Δ_t are bounded in the range [-1, 1] and are

[10] This ensures that all patents in a given year are represented by vectors in the same N_t-dimensional space.

[11] Python code to reproduce this measure is available on the authors' websites.

[12] To ensure z_t and z_{t-1} are in the same space, we modify z_t by adding elements equal to zero for words that are in $t-1$ but not t, and do the analogous procedure for z_{t-1}.

positive if the corresponding word increases in use from year $t-1$ to t and negative if it decreases.[13]

Intuitively, the annual vector Δ_t tracks the appearance, disappearance, and growth of specific technological vocabularies across all patents over time after controlling for the length and number of patents. To build intuition for Δ_t, Table I displays the ten words experiencing the largest increases and decreases in usage in specific years. In 1995, we detect increased use of terms related to genetics, such as "polypeptides", "clones", "recombinant" and "nucleic", following rapid progress in the technology area of genome sequencing. In contrast, the terms "cassette," "ultrasonic," and "tape" are sharply decreasing. In 2005, the most rapidly growing words are related to the internet: "broadband", "click", "configurable", and "telecommunications".

[Insert Table I about here]

Third, we measure the positioning of a given patent j relative to technology cycles (captured by Δ_t) by taking the equally-weighted average of Δ_t over the words used by patent j.[14] We label this new variable "RETech" for "rapidly evolving technology". To implement this calculation, we first define the boolean vector $B_{j,t}$ for each patent j as having elements equal to one for words used by patent j (and zero otherwise). RETech for patent j is then given by

$$\text{RETech}_{j,t} = \left(\frac{B_{j,t}}{B_{j,t} \cdot \mathbf{1}} \cdot \Delta_t \right) \times 100, \tag{3}$$

where $B_{j,t} \cdot \mathbf{1}$ is the number of unique words in patent j.[15] With this definition, patents using words (i.e., non-zero elements in $B_{j,t}$) that are contemporaneously surging across the patent corpus (i.e., larger corresponding elements in Δ_t) receive higher scores for

[13] Normalizing by $(z_t + z_{t-1})$ instead of z_{t-1} ensures that a brand new word in year t has a defined rate of change. In unreported analyses, we show that our results are similar if we build our measure at the quarterly frequency instead of the annual frequency, and if we consider change in word usage over a two year period instead of one.

[14] Considering alternative weighting schemes to compute RETech (e.g., by word usage intensity ($\frac{V_{j,t}}{V_{j,t} \cdot \mathbf{1}}$) instead of binary equal weights ($\frac{B_{j,t}}{B_{j,t} \cdot \mathbf{1}}$)) does not materially change the results of our analyses. We use equal weights in this calculation following Hoberg and Phillips (2016).

[15] Similar to Δ_t, the dimension of $B_{j,t}$ is given by the union of words in all patent in year t and $t-1$.

RETech. Conversely, patents using words that are not growing (or decreasing) in overall usage receive lower scores for RETech.

Continuing with the examples of words from Table I, patents intensely using words such as "polypeptides", "clones", "recombinant" and "nucleic" in 1995 are classified as pertaining to a rapidly evolving area. Their high values for RETech arise from the fact that these words soared in usage in 1995, reflecting rapid breakthroughs in genome sequencing methods, initiating a new technology cycle. Importantly, our methodology classifies patents using the exact same words in 2005 as being in more "stable" technology areas (i.e., lower values of RETech), as genomics methods had matured and the patent vocabulary became stable at that time, reflecting later stages in that technology cycle.

A defining feature of RETech is that it is measurable ex ante, as it does not rely on forward-looking information (e.g., information in future patents such as citations). Moreover, this feature makes RETech potentially valuable in many settings. For example, it could offer value to practitioners and regulators seeking actionable and timely information about startup valuation, job creation, and estimates of economic value.

C Technological Breadth and Similarities

Because the positioning of patents in technology cycles could be correlated with other important technological features, we construct several other text-based measures. First, we compute the technological breadth for each patent. We consider the six major technological fields (f) indicated by the NBER technical classification: chemicals, computer and communication, drugs and medicine, electricity, mechanics, and "other". We count how often each word appears in patents classified into each field in each year. We then tag a word as "specialized" (and associated with field f) in year t if its use in its most prominent field f is more than 150% that of its second most prominent field in year t. Each word is thus classified into one of the six fields or it is deemed to be an "unspecialized" word. For example, "bluetooth" and "wifi" are in the "computer and communication" field, and "acid" and "solvent" are in the "chemicals" field. Finally, we drop unspecialized words and define $w_{j,t,f}$ as the fraction of patent j's specialized words that are classified into each field f. By construction, $w_{j,t,f}$ lies in the [0,1] interval, and they sum to one for

each patent j in year t. We then define technological breadth as one minus technological concentration for each patent:

$$\text{Tech Breadth}_{j,t} = 1 - \sum_{f=1}^{6} w_{j,t,f}^2. \tag{4}$$

Patents have high technological breadth when they draw vocabulary from many fields.

We also separately compute the technological similarity of each patent to the patents of economically linked firms by comparing its vocabulary to that of the patents assigned to firms in three groups: lead innovators, private U.S. firms, and foreign firms. We use cosine similarity measures for parsimony (see Sebastiani (2002)) and ease of interpretation given they are bounded in $[0, 1]$. We define "Lead Innovators" (henceforth LI) as the ten U.S. public firms with the most patent applications in each year. This set, which includes Microsoft and Intel in 2005 and General Electric and Dow Chemical in 1985, varies as the importance of sectors and firms changes. In year t, we identify the set of patents applied for by the LIs over the past three years ($t-2$ to t) and compute the LI vector in year t ($V_{LI,t}$) using the aggregate frequency of word usage across these patents. The similarity of any patent to those of the LIs is:

$$\text{LI Similarity}_{j,t} = \frac{V_{j,t}}{\|V_{j,t}\|} \cdot \frac{V_{LI,t}}{\|V_{LI,t}\|}. \tag{5}$$

We also compute the similarity between the text in each patent j and the overall text of patents assigned to private U.S. firms or to foreign firms, using the patent categorization method described in the Internet Appendix (Section IA.A) to identify patents in each group and form their aggregate vectors.[16]

D Descriptive Statistics

Table II presents descriptive statistics for RETech over the whole sample period.[17] RETech is available for 6,594,248 patents. We first note that the empirical distribution of RETech is highly skewed. The first row indicates that the average value is 1.64, the median is 1.27, and the 75^{th} percentile is 2.34. The observed asymmetry indicates that while the vast

[16] Because these groups contain very large numbers of patents, we aggregate them over just one year, t.
[17] Patent-level variables are winsorized at 1/99% level annually.

majority of patents relate to technology areas that are stable (i.e., low values of RETech), a smaller set of patents appear to pertain to rapidly evolving technology areas (i.e., large values of RETech).

[Insert Table II about here]

Table II also provides statistics for other patent characteristics. Patent breadth is more evenly distributed, indicating less skewness in technological specializations. We also observe some variation in similarity across patents, but the overall levels are low, which is not surprising given the large range and diversity in the vocabulary used across all patents. Statistics for other existing patent variables, namely their originality (Hall, Jaffe, and Trajtenberg (2001)), scope (Lerner (1994)), number of cites, number of claims (Lanjouw and Schankerman (2001)), and economic value (Kogan, Papanikolaou, Seru, and Stoffman (2016), abbreviated as KPSS henceforth) are in line with existing research.[18]

[Insert Table III about here]

Table III presents correlations between different patent-level variables. Several patterns are worth mentioning. First, RETech is only weakly correlated with the other technological attributes of patents (correlations ranging between -0.32 and 0.20), indicating that our new measure is truly capturing a novel aspect of innovation. Second, RETech and breadth are negatively related, indicating that patents in rapidly evolving areas tend to be more specialized. Third, consistent with the idea that patents in rapidly evolving areas are distinct from existing patents, they are overall less similar to other patents, with the exception of patents of lead innovators. Fourth, patents in rapidly evolving areas attract more future citations, and display larger economic value (using the KPSS measure) and scope. Although weak, we nevertheless account for these correlations (excluding the KPSS measure, which is only defined for public firms) in our later cross-sectional analysis

[18]We only use citations received within five years of the grant date. Because we have citation data through 2013, this measure is unlikely to have much truncation bias. As a byproduct, citations per patent is lower in our sample than in studies using future citations over the life of the patent.

of startup exits by systematically controlling for the various technological characteristics of patents.[19]

To illustrate when RETech detects periods of technological acceleration within patented innovation, we report how the aggregate stock of RETech has evolved from 1930 to 2010 across three cuts of the sample.[20] Figure II plots the stock aggregated across all patents (smoothed using a four quarter moving average). Reassuringly, short-term spikes in RETech correspond to well-known waves of technological breakthroughs. For instance, the period around 1950 is known for radical innovation in mechanics, featuring the invention of the transistor, jet engine, and xerography. A second peak occurs in the mid-seventies, corresponding to innovations related to the computer. The last two peaks of rapid technological change appear in the late eighties and mid-nineties, reflecting waves of inventions related to genetics (e.g., methods of recombination) and mass adoption of the Internet.

[Insert Figures II, III and IV about here]

Figure III plots the evolution of RETech across the six broad technology fields defined by the NBER (i.e., Mechanical, Drugs & Medicine, Mechanics, Computers & Communication, Electricity, and Others). Again, we clearly detect well-known historical episodes of technological breakthroughs. Figure IV indicates that the close mapping between aggregate RETech and well-known innovation episodes also holds across four categories of patent assignees (i.e., individuals, private, public, and international firms). Further tests (reported in the Internet Appendix for brevity) indicate that these patterns, and in particular the detected spikes, are not due to shifts in patenting across geographic locations, the composition of assignees and patent attorneys, or the explosion in the number of words used by patents.

[19] Because *Foreign Similarity* and *LI Similarity* are non-trivially correlated (62% and 41%) with *Private Similarity*, in regressions, we orthogonalize *Foreign Similarity* and *LI Similarity* by subtracting *Private Similarity*.

[20] The aggregate stock is the average RETech across all patents applied for in the prior 20 quarters, after applying a 5% quarterly rate of depreciation to each patent. Although we report aggregate patent *stocks*, all patterns remain if we instead report the annual *flow* of RETech across patents, indicating that the patterns are not an artifact of our depreciation assumption.

III Rapidly Evolving Areas and Substitution

Our new measure, RETech, builds on the idea that patented innovation in rapidly evolving technology areas occurs around periods of breakthroughs and is likely to substitute for existing technologies, whereas innovation in stable areas is more likely to complement existing technologies. This section provides evidence examining this proposition.

A Substitution as Revealed by Citation Patterns

Our validation follows Funk and Owen-Smith (2016), who provide direct citation-based formulas to measure the extent to which a given patent has strong substitution or complementarity properties. For example, the approach for substitution examines if a given patent crowds out the prior art it cites in terms of follow-on citations. The authors define "follow-on" patents as all patents that cite either the focal patent or any of the patents the focal patent cited, which we call "predecessors." We follow their approach and we also fix a ten year window to identify the set of follow-on patents for each focal patent. This approach ensures uniform treatment of patents regardless of their vintage.

We thus measure substitution as the number of follow-on patents that cite the focal patent but none of its predecessors, divided by the total number of follow-on patents. When this ratio is high, the focal patent is essentially replacing the prior patents in the evolution of knowledge and thus has strong technological substitute properties. In contrast, we measure complementarity as the number of follow-on patents that simultaneously cite the focal patent and at least one of its predecessors, divided by the total number of follow-on patents. When this ratio is high, the focal patent is cited together with prior patents and thus has strong technological complementarity properties.

[Insert Figure V and Table IV about here]

Figure V plots both substitution and complementarity measures in event-time, contrasted with the RETech of the focal patent. Panel A confirms that the likelihood of technological crowding-out varies strongly and positively with our RETech measure. Patents in the top decile of the distribution of RETech display levels of substitution that are more

than 20% higher than patents in the lowest decile. Panel B reports the reverse for complements: Patents in stable technological areas (low values of RETech) are much more likely to complement their predecessors.

To establish this descriptive finding formally, we regress substitution and complementarity on RETech, including cohort and technology category fixed effects to isolate variation across patents granted in the same year in the same technology area, in addition to assignee type fixed effects (see Internet Appendix IA.A).[21] We cluster standard errors by cohort. Table IV presents the results. The first column reveals a positive and significant coefficient (t-statistic of 9.88), confirming the strong positive relation between RETech and technological substitution. Supporting the limited technological complementarity of patents in rapidly evolving areas, column (2) reports a negative and significant link between RETech and complementarity (t-statistic of -16.72). In columns (3) and (4), we further control for heterogeneity in patents' citation patterns across broad technology fields by including interactions between cohort and technology fixed effects and obtain similar results.

B Breakthrough Inventions

To provide further validation, we ask whether inventions that are recognized by external sources as technological breakthroughs occurred in areas that were rapidly evolving or stable. We specifically consider a collection of 101 breakthrough patents granted between 1930 and 2010 identified by Kelly, Papanikolaou, Seru, and Taddy (2019) based on the USPTO's "Significant Historical Patents of the United States", and several on-line lists. This collection includes for instance inventions such as the "Complex computer" (#2668661) and "DNA modifications" (#4399216), the satellite (#2835548), the laser (#2929922), or PageRank (#6285999). These breakthrough inventions unambiguously shifted the path of technology in their respective areas and created new sustainable markets. Most also crowded out and substituted for existing technologies in their areas.

[Insert Table V about here]

[21] Technology fixed effects are based on two-digit NBER technology codes (Hall, Jaffe, and Trajtenberg (2001)).

We score each of these breakthrough patents based on its percentile rank across various characteristics in its grant year. For instance, a value of 0.95 indicates that the patent is in the top 5% of the respective distribution. We report the scores for each individual patent in the Appendix Table A3. Patents ranking in the top percentile of the RETech's distribution include for instance the neutronic reactor (#2708656), molecular chimeras (#4468464), or the first breast cancer detection method (#5747282). Table V summarizes the results. We find that breakthrough patents mostly occurred when their respective technology areas were in a fluid stage of rapid evolution, as they rank at the 72^{nd} percentile of the RETech's distribution on average, and the 82^{nd} percentile for the median. These percentiles are rather precise as the standard error is just 0.03.

IV RETech and Startup Exits

To further establish the economic relevance of RETech, we study the association between startup exit choices and whether they innovate in technology areas that are rapidly evolving or stable.

A Motivation and Hypotheses

A successful exit is one of the most important events in the life of a startup, and its prospect largely conditions the ex ante incentives and actions of entrepreneurs and their investors. Success is typically achieved when a startup sells shares to outside investors during an initial public offering (IPO) or when it directly sells out to another company. While a large literature investigates the determinants of startups' overall success (see Da Rin, Hellman, and Puri (2013) for a survey), much less is known about the factors driving their specific choice of exit, especially the role of technology-related factors. A better understanding of these factors is important given the key role played by exit opportunities in the development of entrepreneurial activities and innovation (Phillips and Zhdanov (2013) and Kamepalli, Rajan, and Zingales (2020)).

We posit that startup exit choices depend on positioning in technology cycles. We start with a simple rational framework in which a successful startup (i.e., the entrepreneurs and their backers) chooses the exit method that maximizes the value received by equity holders

upon exit, V. We focus on "successful" startups, those for which the exit value V is larger than the "reservation" value of not exiting. The exit value V is the total value conveyed to the startup by outsiders, that is, dispersed investors when startups exit through an IPO or a sale to a strategic buyer. The theory of the firm (Grossman and Hart (1986) and Hart and Moore (1990)) suggests that the startup exit choice depends on the optimal firm boundaries, that is, on whether its exit value is higher as a stand-alone entity operating independently, V_s, or when it is combined with the assets of another entity, V_c. It is optimal for the startup to exit through an IPO when $V_s > V_c$, and by selling out to another entity when $V_c > V_s$.

We propose that a startup's value under each scenario (V_s and V_c) depends on the characteristics of the innovation it develops and its relative position in technology cycles. Specifically, startups innovating in rapidly evolving technology areas should favor exiting by going public because they command higher exit values as independent stand-alone firms relative to selling out. This prediction is motivated by the potential of inventions in rapidly evolving technology areas to substitute and replace existing technologies. If these innovations are in established product markets, startups have the possibility of taking a large share of an existing and proven market from incumbents, while innovations in newly commercialized product markets could give the innovator a first-mover competitive advantage. Exploiting either opportunity depends more on rapid growth and less on strategic assistance and technological complementarity from potential acquirers (Bayar and Chemmanur (2011)). By maintaining independence via an IPO exit, initial equity holders capture more economic rents. These factors predict that $V_s > V_c$ for these startups and, thus, that they should be more likely to exit via IPO.[22]

On the other hand, startups innovating in stable technology areas can achieve a higher exit value when combined with another firm than as a stand-alone entity ($V_c > V_s$). This is because innovation in stable areas tends to complement existing inventions through incremental improvements and synergies with existing technologies (Cassiman and Veugelers

[22]Supporting this hypothesis, Darby and Zucker (2002) show that biotechnology firms go public when their innovations can be successfully commercialized, Chemmanur, He, He, and Nandy (2018) report that manufacturing firms with a defensible market share favor IPOs, and Poulsen and Stegemoller (2008) and Cumming and Macintosh (2003) show that higher growth potential startups tend to exit through IPOs.

(2006)). Hence, the value created through integration with a buyer should exceed the stand-alone value (Higgins and Rodriguez (2006)).[23] Existing studies further suggest that complementary technologies generate synergies by facilitating follow-on innovation (Holmstrom and Roberts (1998)) and improving technology coordination (Hart and Holmstrom (2010)).[24] In addition, when the target's technology is overlapping and complementary, synergies can be created because the buyer's resources can foster product market success. Overall, the synergistic value created from asset combination can outweigh the costs of sharing future economic rents with the buyer for startups innovating in stable technology areas. Hence, such startups should be more likely to exit by selling out.

Although the existing literature provides a solid foundation for our proposed link between the substitution or complementary potential of startups' technologies and their choice of exits, other hypotheses might be relevant. We note three examples. First, the potential for technological substitution could also spur "killer" acquisitions by threatened incumbents (e.g., Cunningham, Ederer, and Ma (2021)), predicting a positive association between high RETech startups and sell-outs. While evidence that killer acquisitions exist is compelling, they appear relatively uncommon (e.g., 6% of acquisitions in the pharmaceutical industry appear to have this motive according to Cunningham, Ederer, and Ma (2021)), which might explain why we do not detect this channel in our sample. Second, innovation in rapidly evolving areas might also be related to observed exits because it entails higher levels of risk. Yet, analyses based on startup failures and various risk metrics (unreported for brevity) suggest that this is not the case.[25] Finally, startups innovating in

[23]While we focus on the associations between patents' characteristics and acquisition of whole companies (similar to Bena and Li (2014) or Fresard, Hoberg, and Phillips (2020)), value from technology complementarity could also be created via the "market for ideas", through patent sales or licensing agreements. However, we believe that focusing on acquisitions represents a useful setting to validate the economic relevance of RETech because of the existence of various frictions in this market (e.g., the need for employees' skills to integrate technologies) and because the startups we focus on have to exit promptly due to the limited lives of most VC funds.

[24]Innovation in stable technology areas could also attract higher valuations in a sell-out because buyers might have better information about potential synergistic gains obtained from combining well-understood technologies (Kaplan (2000) or Hirshleifer, Hsu, and Li (2017)).

[25]For instance, Nanda and Rhodes-Kropf (2013) indicates that, in hot markets, innovative startups are more risky, which affects their outcomes (e.g., higher rate of failures). Although the logic of our validation test relies on the distinct exit choices of successful startups, we recognize the possibility that innovation in rapidly evolving areas could be inherently more risky, but fail to find consistent evidence for this channel in our sample.

stable technology areas whose patents are complementary to existing firms might choose to sell or license the patents rather than sell the entire firm. To the extent that this channel matters, we note that it would reduce our estimated relation between RETech and sell-outs towards zero.[26] Our finding of a significant coefficient therefore indicates that patent transactions can only be part of the story.

B Panel of VC-Backed Startups

To test these predictions, we use data on VC-backed U.S. startups from Thomson Reuters's VentureXpert (Kaplan, Stromberg, and Sensoy (2002)), which contains detailed information including the dates of financing rounds and their ultimate exit (e.g., IPO, acquisition, or failure). We focus on the period 1980-2010 and restrict attention to startups that are granted at least one patent during the sample period.[27] To link patents to startups, we follow Bernstein, Giroud, and Townsend (2016) and develop a fuzzy name matching algorithm (see Section IA.B in the Internet Appendix) mapping the names of patents (initial) assignees to that of startups.[28]

A startup enters our startup-quarter sample in the quarter it is founded (based on founding dates in VentureXpert) and exits the sample when its final exit is observed on the "resolve date" from VentureXpert. Startups still active in November 2017 are unresolved.[29] We exclude startups if their founding date is missing or is later than the resolve date. Our final sample contains 347,929 startup-quarter observations, corresponding to 9,167 unique startups and 94,703 patent applications.

To obtain the positioning of a startup in the technology cycle in a given quarter, we

[26]Using data from the USPTOs Patent Assignment Dataset, we find that startups participate in the patent market on a very infrequent basis: Only 1.1% of startup-quarters have patent sales.

[27]This time period is where high quality data linking patents to public firms (which ends in 2010) overlaps with VentureXpert (which starts in 1980).

[28]Lerner and Seru (2017) note that bias can occur in matching patent assignments to firms because patents can be assigned to subsidiaries with different names than their parent corporations. However, this issue is limited in our sample as startups have simpler corporate structures. In addition, by focusing on patents' initial assignees, we might miss patents that are assigned to startups later on (e.g., patents assigned to inventors that start companies or those purchased by the startup). This is a data limitation that is common to our study and many other studies, and this is also an area that can greatly benefit from future research.

[29]Ewens and Farre-Mensa (2020) note that unresolved firms can result from stale data collection, and we code firms as failed if seven years pass since their last funding round.

aggregate the RETech of its existing patents. Conceptually, we think of a startup as a basket of technologies, and because it is a startup, this basket is likely to be quite focused (unlike large public firms that can have sprawling technology categories covered). Our framework thus assumes that any observed patented component is indicative of a startup's overall innovation portfolio. Other components might still be under development or are trade secrets, but they are likely in the same technological areas as the observed patent.

We thus obtain the value of RETech for a given startup i in a given quarter q by taking the average RETech across all of startup i's patent applications over the prior 20 quarters (from quarter $q-19$ to q), while applying a 5% quarterly rate of depreciation.[30] Because a startup's innovation is typically intermittent and unevenly spread over time, this method of aggregation implicitly captures startup i's' "stock" of RETech (measured in quarter q) scaled by the number of patents so that its RETech stock does not reflect the size of its patent portfolio.[31] Note that, by construction, RETech is equal to zero when startups have no patent applications over the prior 20 quarters (thus we control for a zero-patent dummy). We perform a similar aggregation for all the other patent-based variables used in our analysis.[32] We define exit variables (IPO or sell-out) as binary variables equal to one if startup i experiences a given exit in quarter q. Detailed variable constructions are given in Table A1.

Although our sample does not include non-VC backed startups, we note that VC-backed startups are a useful laboratory to study the link between the type of innovation developed by startups and their exits. First, these firms account for a large share of the IPO market (Ritter (2019)) and the production of innovation (Gornall and Strebulaev (2015)). Second, we show in the Internet Appendix that their IPO and acquisition rates over the last thirty years are comparable to the economy-wide patterns. Third, VC-backed

[30]The formal equation for this aggregation is in Table A1. We use the 5% quarterly rate to match the 15-20% annual depreciation rate on citation stocks common in the literature. Our findings are robust to assuming zero depreciation. Nevertheless, we believe that depreciation is economically sensible for our new technology characteristics as well. For example, it is intuitive to expect that if two patents used the same words five years apart, the latter will have a lower RETech.

[31]We show in the Internet Appendix (Table IA.1) that our main results are robust to different ways of aggregating patent-level information at the level of startups.

[32]All underlying patent-level variables are winsorized at 1/99% level annually before we calculate startup-level stocks.

startups generally exit promptly due to the limited lives of most venture capital funds, allowing us to study them without long waiting periods to resolution.

[Insert Table VI about here]

Overall, Table VI indicates that startup-quarter statistics are similar to that of patents (see Table II), indicating that the technological characteristics of VC-backed startups are roughly representative of those in the economy at large. Relevant for our regression analysis, Table VI further indicates that the quarterly IPO rate (i.e. the number of IPOs in a quarter divided by the number of active startups in that quarter) is 0.42 percentage points, and the quarterly sell-out rate is 0.73 percentage points.[33]

C Startup RETech and Exits

We first compare the empirical distribution of startups' RETech depending on their chosen exit (measured during the quarter of their exit). Figure VI confirms that they differ. In particular, we observe that startups exiting by going public exhibit higher values of RETech compared to those exiting by selling out. The difference is large (and statistically significant), as the average (median) value of RETech for IPO exit is 1.55 (1.24), compared to 1.10 (0.87) for sell-outs. From a related perspective, Figure VII reveals that the fraction of IPO exits in our sample is substantially higher for startups displaying high levels of RETech. While only 20% of low RETech startups exit by going public, close to 50% of high RETech startups choose to exit through IPOs.

[Insert Figure VI and Figure VII about here]

Although in line with our hypothesis, drawing conclusions from these descriptive findings is difficult since they ignore (i) the possible dependence between startups' exit options, and (ii) the possible influence of variables correlated with both startups' RETech and exit choices (e.g., startups' quality or other technological features). To extend these

[33] We report additional information about the life cycle of sample firms in the Appendix in Table A2. Relative to the founding date, the average firm applies for its first patent in 4.42 years and receives its first VC round in 5.29 years. IPOs occur faster (9.41 years) than sell-outs (11.23 years) on average.

univariate results, we use our startup-quarter panel (see Section IV.B) to estimate, in a multivariate setting, whether innovating in rapidly evolving or stable areas is associated with startups' exit, after controlling for other (observed and unobserved) determinants of exits and accounting for the dependence among exit choices.

To do so, we use the competing risks model of Fine and Gray (1999) to explicitly account for the "risk" of observing a particular startup's exit (i.e., an event) in a given quarter. Indeed, in each quarter, startups could exit the sample through an IPO, a sell-out, or a failure. Since only one of these event can occur for a given startup, we view these events as "competing events", in the sense that the occurrence of one event precludes the occurrence of the others. These are competing risks because the probability of each competing event is affected by the other competing events. In our estimations, the dependent variable (i.e., the distinct competing events) is either an IPO or a sell-out measured in quarter q.[34] The model estimates the marginal probability for each competing event as a function of a set of covariates (i.e., explanatory variables), without assuming the independence of competing events (nevertheless we show later that our results are robust to simpler models).[35] We cluster standard errors by startup.

All explanatory variables are measured with a lag (i.e., in quarter $q-1$).Besides our key variable of interest, RETech, we consider a host of startup and market characteristics that may correlate with RETech and also be related to a startup's exit choice. In particular, to control for other technological specificities of startups that might drive exit choices, we include their technological breath, our three text-based measures of patent similarity, patents' citations, originality, scope, and their number of claims (all aggregated similarly as RETech, see above), as well the number of granted patents in the prior 20 quarters and a dummy variable identifying the absence of patent applications during the prior 20

[34]We allow for failure as a third competing event, but do not report models focusing on failure as an outcome, as failure is not uniformly coded in VentureXpert on a timely basis (see Ewens and Farre-Mensa (2020)).

[35]Similar to a Cox hazard model, a competing risks hazard model explicitly models the "risk" of choosing a particular exit at a particular time. However, unlike a Cox hazard model, it accounts for multiple potential hazards that are mutually exclusive. See Avdjiev, Bogdanova, Bolton, Jiang, and Kartasheva (Forthcoming) for a recent study using competing risk models in finance. We use the function *stcrreg* in Stata to estimate the models.

quarters.[36]

Furthermore, we include several variables to control for potential differences in startups' quality, which could arguably be related to RETech and drive their exit. For instance, high quality startups favoring IPO exits might display higher RETech, not because of the genuine attributes of their technology, but because of the selective wording chosen by patent lawyers (as startup quality likely affects their choice of lawyers). Because VC funds progressively learn about the quality of startups over time (e.g., by staging their investments), we consider startups' past VC financing pattern as a proxy for their quality (as perceived by the VCs). We thus include the cumulative VC funding received from startups' founding date until quarter q, as well as a binary variable identifying the absence of any VC funding during the prior 20 quarters. Interestingly, we report in the Appendix (see Table A4) that startups innovating in rapidly evolving areas receive significantly more frequent and larger amounts of VC funding, which confirms the importance of controlling for differences in VC funding in our main specification.[37] Because better VCs are usually able to invest in higher quality startups, we further proxy for startup quality using the quality and reputation of its VCs, measured using VCs' market shares of investment capital (Krishnan and Masulis (2012) or Bengtsson and Sensoy (2011)).

Finally, following the literature on IPOs and acquisitions (Lowry (2003), Pastor and Veronesi (2005) or Betton, Eckbo, and Thorburn (2007)), we also control for aggregate stock market activity that could coincide with startups' innovation choices and exits (e.g., startups developing more innovation in rapidly evolving technologies and exiting via IPOs in hot markets). Specifically, we include the aggregate market-to-book ratio (using all firms in Compustat), the market return (from Ken French's data library), as well as dummy variable capturing the last quarter of the year to control for exit cyclicality.

[36] We include this variable to distinguish situations where RETech is zero because a startup does not apply for any patent during the last five years from situations where RETech is really equal to zero.

[37] Note that the positive relation between startups' RETech and VC funding could indicate that startups in rapidly evolving areas may seek to create a first mover advantage or carve out market share from an incumbent, and these growth strategies require larger investments than those undertaken by startups with incremental and complementary technologies. For example, a startup with stable and complementary technology might need less financing to invest in corporate infrastructure, as potential acquirers will likely provide this infrastructure.

[Insert Table VII about here]

The first two columns of Table VII (Panel A) display the results. To ease interpretation, we report standardized exponentiated coefficients from which we subtract one, so that the tabulated coefficients indicate the effect of a one standard deviation increase in a given variable on the incidence of IPO or sell-outs *relative* to their baseline incidence across startups (i.e., their value when all variables are at their sample mean). Mirroring the univariate results, column (1) indicates a strong positive relationship between a startups' RETech and the likelihood of exiting via IPO in the next quarter. The point estimate is 0.247 with a t-statistic of 12.26. In contrast, column (2) shows that sell-out exits are negatively related to RETech with a coefficient of -0.143 and a t-statistic of -5.53. Confirming the economic relevance of our new measure, we observe a stark difference in exit choice between high and low RETech startups.

These results are not only statistically significant, but the coefficients also indicate economically large magnitudes: a one standard deviation increase in RETech is associated with a *proportional* 24.7% increase in quarterly IPO incidence. Given the baseline quarterly IPO incidence of 0.42% (see Table II), our estimates indicate that a one standard deviation increase in RETech is associated with an increase in startups IPO incidence from a baseline of 0.42% to 0.52%. A similar increase in RETech is associated with a *proportional* 14.3% decrease in the incidence of sell-outs (i.e., from a baseline incidence of 0.73% to 0.62%). Across specifications, the coefficients on startups' RETech display an economic magnitude that is similar (or slightly larger) to that of other technology variables (e.g., number of patents, citations, originality, or scope).

Overall, the coefficients on the control variables display the expected signs. For instance, startups' technological breadth is positively related to IPO incidence and negatively related to sell-out incidence, confirming that high breadth technologies are less redeployable and have lower potential synergies (Bena and Li (2014)). We also find that startups whose patents are more similar to those of other private firms are significantly less likely to exit through sell-outs (t-statistic of -6.72) and are marginally more likely to go public (t-statistic of 1.83). These results are consistent with the negative link between

product market similarity and acquisitions for public firms documented in Hoberg and Phillips (2010). Startups holding patents that resemble those of successful lead innovators are also significantly more likely to go public (t-statistic of 4.26). Startups with a non-zero number of patents in the most recent years are more likely to sell-out than go public. The total patents variable offsets this effect, as startups with higher overall patent counts are associated with a shift towards IPOs.[38] More future citations (originality) are positively (negatively) associated with exits via both IPOs and sell-outs.[39]

In addition, higher quality startups, either measured using their financing patterns or the market share of their VC, display larger incidence of exits. These results underscore that even startups exhibiting similar pre-exit growth financing trajectories choose distinct exit forms depending on whether they innovate in rapidly evolving or stable technology areas. Finally, and consistent with earlier research, we find that startups are more likely to exit via IPO after periods of strong stock market performance.

D Unobserved Heterogeneity Across Startups

In the rest of Table VII, we replace our baseline competing risk model with linear probability models for each exit type. Although linear models ignore potential dependence across different types of exits (i.e., competing risks), they allow us to include a wide array of fixed effects to limit the potential role of unobserved startups' characteristics (i.e., not capture by the baseline controls) on our conclusions. Specifically, we saturate our specifications with fixed effects for year, startups' industry (using VentureXpert's major industry group), technology areas, state of location, age (based on their founding year), and cohort (i.e., year of first VC funding).[40] This fixed effect structure allows us to specifically study the role played by RETech on exits in a given year among startups of the same age and funding vintage, located in the same state and active in the same industry and technology area. Such a stringent approach allows us to absorb any systematic time-

[38] For example, the estimates in column (3) indicate that, compared to a startup with no patents, a startup with one patent is 0.40 p.p. less likely to go public while those with 25 or more patents are more likely to go public.

[39] With the exception of patent citations, which we include to be consistent with the literature, all variables are ex ante measurable. Our results are fully robust to excluding this variable from the model.

[40] Technology fixed effects are based on the most prevalent NBER technology category used within a firm's patent portfolio (see Lerner and Seru (2017)).

invariant unobserved variation in RETech and exits related to these various dimensions (e.g., startups in California or active in biotech favoring rapidly evolving technologies and independently preferring IPO exits).

Columns (3) and (4) of Table VII show that our results continue to hold for linear probability models. The statistical inferences and economic magnitudes are also very similar to those for the competing risks model in the first two columns. For instance, the coefficient of 0.073 in column (3) indicates that a one standard deviation increase in RETech is associated with a 0.073 percentage point larger IPO incidence, representing a 17.4% increase compared to the baseline incidence of 0.42% (from 0.42% to 0.49%).

To further limit the possible impact of startups' heterogeneous quality, we consider a separate specification where we limit attention to startups exiting successfully (i.e., the subset of quality startups for which the exit value V surpasses the reservation value of not exiting). That is, in column (5) of Table VII, we keep the last observation for startups that go public or sell-out, and estimate a linear model where the dependent variable is 100 for startups exiting via IPO and zero if they sell out.[41] The estimated coefficient on RETech is positive, statistically significant (t-statistic of 2.03), and economically meaningful: conditional on exiting, a one standard deviation increase in RETech is associated with a 2.18 percentage point higher likelihood that exit is via IPO instead of a sell-out. Relative to the baseline IPO incidence of 36.1% in this sample, this implies that a one standard deviation shift in RETech is associated with a 6.0% increase in the IPO incidence relative to the baseline incidence.

E Robustness

We consider a number of important robustness checks. First, we alter the baseline econometric specification and consider logistic and multinomial (logistic) regressions in Panel B of Table VII. For brevity, we only report the coefficients on RETech in Panels B. Our results are similar to those obtained with our competing risk model.[42]

[41] Standard errors for this model are clustered by observation year.

[42] We display exponentiated coefficients minus one (i.e. odds ratios minus one), so that interpretation is comparable to the competing risk models.

We then examine whether the results are robust to considering longer exit horizons. To do so, we increase the measurement window for exit incidence from one quarter (i.e., our baseline) to five years using increments of one year and focus on linear probability models including all fixed effects, as described above. For brevity, we only report the coefficients on RETech. Panel A shows that the positive relationship between RETech and IPO exits remains strong at all horizons, and Panel B shows that the negative relation between RETech and sell-outs persists for one year and fades after two.

[Insert Figure VIII and Table VIII about here]

In Panel C of Table VIII, we examine whether startups' RETech is related to their general propensity to exit (or remain private). The dependent variable is a dummy that equals one if a startups stays private next period, with horizons again ranging from one quarter to five years. The results show that startups innovating in rapidly evolving technological areas tend to exit more quickly on average (except at the quarterly horizon), and hence remain private for shorter periods relative to startups in stable areas. Figure VIII, which plots the average time-to-exit (i.e., quarters since the first VC funding round) across exit types and quintiles of RETech, reveals that the faster exit of high RETech startups is entirely driven by faster IPO exits by these startups.

[Insert Table IX about here]

Table IX reports a host of additional robustness tests. In test 1, we include VC fixed effects (based on the lead VC of the most recent funding round) to assess the relation between startups' RETech and exits within lead VCs' portfolios (and hence control for possible VCs' exit preferences). Our results are barely affected. Our results also hold in test 2 where we add industry-year fixed effects to control for the potential effects of time-varying industry factors (such as valuation booms and busts). Our results continue to hold when standard errors are clustered by year, startup technology, or startup cohort (tests 3, 4, and 5). In test 6 we exclude (forward looking) citations as a control, and in test 7, the only control is for the size of a startup's patent portfolio. Test 8 includes access to

bank financing and patent generality as additional controls.[43] Following Maats, Metrick, Yasuda, Hinkes, and Vershovski (2011) who note that some sell-outs are liquidations rather than successful exits, Test 9 isolates sell-outs that are clearly successful exits (larger than $25 million in 2009 dollars) and Test 10 drops sell-outs to financial acquirers to focus on acquisitions where technological synergies are likely more salient. Tests 11 and 12 show that our results are robust in both halves of the sample, ensuring that our results are not driven by either (A) truncation bias from patents not yet granted (Lerner and Seru (2017)) or (B) startups with "unresolved" status. Finally, in Test 13, we reconstruct the sample and variables at an annual frequency and find similar results.[44]

F Established and New Technology Areas

Our findings are consistent with the hypothesis that startups innovating in rapidly evolving areas favor IPO exits because their innovation can substitute existing technologies, allowing startups to create independent market presence. To further assess the validity of this mechanism, we examine if the predictive role of RETech for startup exits is stronger when the economic benefits of technology substitution are larger. We propose that these benefits are larger when technology is rapidly evolving in *established* areas (i.e., existing areas experiencing new waves of innovation) as opposed to new areas, because startups offering technological substitution in established areas can capitalize on existing markets that are larger and proven.

To identify whether startups are innovating in rapidly evolving areas that either established or new, we decompose RETech into two components based on the age of each word in each patent relative to when each word first appears in the patent corpus (based on description sections). We define words as old, and pertaining to established areas, if they have been used in the patent corpus for at least ten years. We define patents in

[43] Our control for access bank financing equals one if the startup takes out a loan in $t-1$ per DealScan. We use adjusted generality, as defined in Hall, Jaffe, and Trajtenberg (2001), which assesses how broadly a patent is cited in the future. This is converted to startup-quarter from patent-level like *RETech*.

[44] For tests at the annual frequency, independent variables are lagged one year, technology stocks are constructed with 20% annual depreciation, and the models do not include the aggregate quarterly variables *Log(MTB), MKT Return*, or *Q4*.

rapidly evolving established technology areas as:

$$\text{RETech(Established)}_{j,t} = \left(\frac{B_{j,t}^{10+} \cdot \Delta_t^{10+}}{B_{j,t}^{10+} \cdot \mathbf{1}} \right) \times 100 \tag{6}$$

where $B_{j,t}^{10+}$ and Δ_t^{10+} are the vectors $B_{i,t}$ and Δ_t defined in Section II.B, except we remove elements relating to words less than ten years old within the patent corpus at time t. By construction, this measure is not influenced by young or new words. For example, patents containing the word "internet" in 1993 will not necessarily display high values of RETech(Established) because the new term "internet" is not part of the established vocabulary. Instead, patents will display high values when they use older words whose usage suddenly surges in volume across all patent applications in a given year, thus belonging to "second (or later) waves" of rapid technology change within a specific established area.[45] Analogously, we identify rapidly evolving patents in new areas as

$$\text{RETech(New)}_{j,t} = \left(\frac{B_{j,t}^{<10} \cdot \Delta_t^{<10}}{B_{j,t}^{<10} \cdot \mathbf{1}} \right) \times 100 \tag{7}$$

where $B_{j,t}^{<10}$ and $\Delta_t^{<10}$ are the vectors $B_{i,t}$ and Δ_t defined in Section II.B, including only elements relating to words less than ten years old within the patent corpus at time t.[46]

[Insert Table X about here]

To verify if the role of rapidly evolving technologies in startups' exit is stronger when the economic benefits of technology substitution are larger, we aggregate both components to startup-quarter variables (as we do for all patent-level variables) and include them in our specifications (replacing RETech). Table X presents the results. Confirming our conjecture, columns (1) and (3) reveal that startups innovating in rapidly evolving established areas are significantly more likely to opt for IPO exits. This is not the case,

[45] An example is patent #7,663,607 for multipoint touchscreens, which was granted to Apple in 2010. This patent introduced new ways to combine existing technologies at a point in time when cell phones, display technology, and user-interfaces were the focus of a wave of rapidly expanding patenting activity.

[46] Apple's aforementioned touchscreen patent resembles the former; it scores in the 84th percentile of the former, but in the 2nd percentile of the latter. Conversely, the earliest patents in a sequence of breakthrough patents governing co-transformation (a method of altering multiple genes) have high levels of RETech(New). Key patents on co-transformation in later years subsequently shifted towards higher levels of RETech(Established). This trend emerges clearly in other technology spaces we examined including semiconductors.

however, for startups innovating in new areas that are evolving rapidly, as the estimated coefficients on RETech(New) are statistically insignificant. These results support the view that the gains to rapid evolution are likely greatest in more established markets, where demand is proven and is likely to be larger in scope. In columns (2) and (4), we observe that the probability of exiting through sell-outs is higher for startups innovating in stable areas, established or new. This suggests that the synergistic benefits from innovating in stable areas are more balanced across established and new areas.

G Ancillary Results Based on Exiting Startups

To further assess the likely economic mechanisms generating our results, we perform three ancillary tests based on sub-samples of startups exiting via sell-out or IPO, whose exits provides us with additional data (unavailable for the whole sample).

First, for a subset of 854 startups selling out to public firms possessing patents (representing about 25% of all sell-outs in our sample), we compare their RETech to that of their acquirer. Indeed, if startups innovating in stable technology areas favor selling out because their innovation tends to complement existing inventions (as we conjecture), we should observe these startups being acquired by companies also active in stable technology areas. To verify this prediction, we compute the RETech of public acquirers by aggregating their patents following the same aggregation procedure we use for startups (see Section II.B). We then divide both startups and acquirers into 25 RETech bins and plot in Figure IX the number of acquisitions in each of the 625 pairs of bins. We find a positive relationship between the RETech of acquirers and selling startups, suggesting a positive assortive matching based on RETech. Consistent with our hypothesis, Figure IX reveals that the large majority of sell-outs are consistent with an assortive match between startups and acquirers innovating in stable technology areas, where the combination of complementary technologies is likely most valuable.

[Insert Figure IX and Table XI about here]

Relatedly, we analyze whether the stock market reaction to acquisition announcements, a commonly-used measure for the expected value of synergies (Betton, Eckbo, and

Thorburn (2007)), varies with the selling startup's RETech. We obtain the cumulative abnormal returns (CAR) for acquirers from a Fama-French three factor model estimated over 100 days prior the announcement, and consider returns in event windows of one and three days around the event. In both windows, columns (1) and (2) of Table XI show a negative and significant relationship between acquirers' returns and startups' RETech (measured as of the quarter of the acquisition announcement). These results support the idea that buying startups innovating in stable technology areas likely offers larger synergies to public buyers.

Our next test focuses on a sub-sample of startups exiting via IPO (1,167 startups), and we study how their capital needs relate to their RETech. The intuition for this test is that, in order to capitalize on their substituting technologies to build independent market positions, startups innovating in more rapidly evolving areas should seek to create a first mover advantage or carve out market share from incumbents, and these growth strategies require more and faster capital. We thus posit that these startups need to raise more capital and need to do so faster. Interpreted in light of Clementi (2002) and Chemmanur and He (2011), the required growth will also accelerate the firms progress toward a level of efficiency that would trigger an IPO exit. We assess this prediction in several ways. First, Figure VIII displays the average time-to-exit (since the first funding round) across groups of startups based on their RETech. We find that the fastest exits are found among high RETech startups going public. This group exits 4.5 quarters earlier than low RETech startups also going public, and five quarters faster than startups that sell-out (irrespective of their RETech).

Further consistent with startups innovating in rapidly evolving areas having a higher need for capital, columns (3) and (4) of Table XI show a positive and significant relationship between startups' ex ante pre-IPO RETech and the fraction of primary shares in the IPO (primary shares are those raising new capital rather than cash-outs by pre-IPO shareholders), as well as the likelihood that all IPO shares are primary shares. In a complementary test, we focus on the relationship between startups' RETech and their financing trajectory using our sample of pre-exit startups. We examine whether a startup

receives a new VC round, the dollar amount raised from VCs, and whether the startup received a new syndicated bank loan in a given quarter. In particular, we regress these events on startups' ex ante RETech measured in the prior quarter as well as control variables (similar to our main specifications). Appendix Table A4 indicates that the coefficients on RETech are uniformly positive and significant, confirming that high RETech startups have a higher need for capital and grow faster.

V Conclusion

We develop a new methodology to measure the positioning of firms in technology cycles based on the text of patents to identify whether patented innovation pertains to a technological area that is rapidly evolving or stable. Our measure builds on the idea that innovation in rapidly evolving technology areas (i.e., following breakthrough periods) tends to substitute for existing technologies, whereas innovation in stable areas tends to complement existing technologies. We show that our text-based measure provides a novel characterization of technological innovation that is distinct from existing measures and consistent with our interpretation.

We further demonstrate the economic relevance of our measure by studying how the distinction between innovation in rapidly evolving or stable technology areas relates to startups' modes of exit (i.e., going public or selling out to another company). We find that startups innovating in rapidly evolving areas are significantly more likely to exit via an IPO and are less likely to sell-out. Our results overall suggest that such differential exits might arise because innovation in rapidly evolving technology areas substitutes for existing technologies, offering startups a path to an independent market presence and high stand-alone valuations. In contrast, innovation in stable areas tends to complement existing technologies, and startups favoring these technologies tend to exit by selling out.

Our text-based measures of technological traits are novel and available over many decades. Because they can be computed as soon as a patent's text is made public, they are potentially valuable in many other settings, especially those requiring predictive models or time-sensitive information. For example, they could provide actionable, policy-relevant

information about the health of innovation sector-by-sector. Researchers might also use them to study the performance of VC funds, portfolio selection, technology diffusion, mergers, and industrial organization.

References

Abernathy, William J., and James M. Utterback, 1978, Patterns of industrial innovation, Technology Review 80, 41–47.

Avdjiev, Stefan, Bilyana Bogdanova, Patrick Bolton, Wei Jiang, and Anastasia Kartasheva, Forthcoming, Coco issuance and bank fragility, Journal of Financial Economics.

Balsmeier, Benjamin, Mohamad Assaf, Tyler Chesebro, Gabe Fierro, Kevin Johnson, Scott Johnson, Guan Cheng Li, Sonja Luck, Doug O'Reagan, Bill Yeh, Guangzheng Zang, and Lee Fleming, 2018, Machine learning and natural language processing on the patent corpus: Data, tools, and new measures, Journal of Economics and Management Strategy pp. 535–553.

Bayar, Onur, and Thomas Chemmanur, 2011, IPOs versus acquisitions and the valuation premium puzzle: A theory of exit choice by entrepreneurs and venture capitalists, Journal of Financial and Quantitative Analysis pp. 1755–1793.

Bena, Jan, and Kai Li, 2014, Corporate innovations and mergers and acquisitions, Journal of Finance 69, 1923–1960.

Bena, Jan, Hernan Ortiz-Molina, and Elena Simintzi, 2020, Shielding firm value: Employment protection and process innovation, Working Paper.

Bena, Jan, and Elena Simintzi, 2019, Machines could not compete with chinese labor: Evidence from us firms innovation, Working Paper.

Bengtsson, Ola, and Berk Sensoy, 2011, Investor abilities and financial contracting: Evidence from venture capital, Journal of Financial Economics 20, 477–502.

Bernstein, Shai, Xavier Giroud, and Richard Townsend, 2016, The impact of venture capital monitoring, Journal of Finance pp. 1591–1622.

Betton, Sandra, Espen Eckbo, and Karin Thorburn, 2007, Corporate takeovers, forthcoming Handbook of Corporate Finance: Empirical Corporate Finance: Elsevier/North-Holland.

Callander, Steven, 2011, Searching and learning by trial and error, American Economic Review 101, 2277–2308.

Cassiman, Bruno, and Reinhilde Veugelers, 2006, In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition, Management Science pp. 68–82.

Chemmanur, Thomas, and Jie He, 2011, Ipo waves, product market competition, and the going public decision: Theory and evidence, Journal of Financial Economics 101, 382–412.

——— , Shan He, and Debrashi Nandy, 2018, Product market characteristics and the choice between IPOs and acquisitions, Journal of Financial and Quantitative Analysis pp. 681–721.

Chemmanur, Thomas, Jie He, Xiao Ren, and Tao Shu, 2020, The disappearing IPO puzzle: New insights from proprietary U.S. Census data on private firms, Working paper.

Clementi, Gian Luca, 2002, IPOs and the growth of firms, New York University working paper.

Cockburn, Iain, and Megan MacGarvie, 2009, Patents, thickets and the financing of early-stage firms: Evidence from the software industry, Journal of Economics and Management Strategy 618, 729773.

Cumming, Douglas, and Jeffrey Macintosh, 2003, Venture-capital exits in Canada and the United States, The University of Toronto Law Journal pp. 101–199.

Cunningham, Colleen, Florian Ederer, and Song Ma, 2021, Killer acquisitions, Journal of Political Economy 129, 649–702.

Da Rin, Marco, Thomas Hellman, and Manju Puri, 2013, A survey of venture capital research, Handbook of the Economics of Finance. G. Constantinindes, M. Harris, and R. Stulz (eds).

Darby, Michael, and Lynne Zucker, 2002, Going public when you can in biotechnology, NBER Working Paper.

Ewens, Michael, and Joan Farre-Mensa, 2020, The deregulation of the private equity markets and the decline in IPOs, Review of Financial Studies 33, 54635509.

Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist, 2019, What is a patent worth? Evidence from the U.S. patent lottery, Journal of Finance 75, 639–682.

Fine, Jason, and Robert Gray, 1999, A proportional hazards model for the subdistribution of a competing risk, Journal of the American Statistical Association pp. 496–509.

Fresard, Laurent, Gerard Hoberg, and Gordon Phillips, 2020, Innovation activities and integration through vertical acquisitions, Review of Financial Studies 33, 29372976.

Funk, Russell J, and Jason Owen-Smith, 2016, A dynamic network measure of technological change, Management Science 63, 791–817.

Gaule, Patrick, 2018, Patents and the success of venture-backed startups: Using examiner assignment to estimate causale effects, Journal of Industrial Economics 66, 350–376.

Gornall, Will, and Ilya Strebulaev, 2015, The economic impact of venture capital: Evidence from public companies, Working Paper.

Grossman, Sanford J., and Oliver D. Hart, 1986, The cost and benefits of ownership: A theory of vertical and lateral integration, Journal of Political Economy 94, 691–719.

Guzman, Jorge, and Scott Stern, 2015, Where is Silicon Valley?, Science pp. 606–609.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, 2001, The NBER patent citation data file: Lessons, insights and methodological tools, No. w8498 National Bureau of Economic Research.

Hanley, Kathleen, and Gerard Hoberg, 2010, The information content of IPO prospectuses, Review of Financial Studies 23, 2821–2864.

Hart, Oliver, and Bengt Holmstrom, 2010, A theory of firm scope, Quarterly Journal of Economics pp. 483–513.

Hart, Oliver, and John Moore, 1990, Property rights and the nature of the firm, Journal of Political Economy 98, 1119–1158.

Higgins, Matthew, and Daniel Rodriguez, 2006, The outsourcing of R&D through acquisitions in the pharmaceutical industry, Journal of Financial Economics pp. 351–383.

Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li, 2017, Innovative originality, profitability, and stock returns, Review of Financial Studies 31, 2553–2605.

Hoberg, Gerard, and Gordon Phillips, 2010, Product market synergies in mergers and acquisitions: A text based analysis, Review of Financial Studies 23, 3773–3811.

——— , 2016, Text-based network industry classifications and endogenous product differentiation, Journal of Political Economy 124, 1423–1465.

Holmstrom, Bengt, and John Roberts, 1998, The boundaries of the firm revisited, Journal of Economic Perspective pp. 73–94.

Kamepalli, Sai Krishna, Raghuram Rajan, and Luigi Zingales, 2020, Kill zone, Working Paper.

Kaplan, Steven, 2000, Mergers and productivity, Bibliovault OAI Repository, the University of Chicago Press.

——— , Per Stromberg, and Berk Sensoy, 2002, How well do venture capital databases reflects actual investments?, Working Paper.

Kaplan, Steven N, and Luigi Zingales, 1997, Do investment-cash flow sensitivities provide useful measures of financing constraints?, Quarterly Journal of Economics 112, 169–215.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, 2019, Measuring technological innovation over the long run, Working Paper.

Klepper, Steven, 1996, Entry, exit, growth, and innovation over the product life cycle, American Economic Review 86, 562–583.

Kogan, Leonid, Dimtris Papanikolaou, Amit Seru, and Noah Stoffman, 2016, Technological innovation, resource allocation and growth, Quarterly Journal of Economics pp. 665–712.

Krishnan, C.N.V, and Ronald Masulis, 2012, Venture capital reputation, The Oxford Handbook of Venture Capital . D. Cumming (ed).

Lanjouw, Jean, and Mark Schankerman, 2001, Tcharacteristics of patent litigation: A window on competition, Rand Journal of Economics 32, 129–151.

Lerner, Josh, 1994, The importance of patent scope: An empirical analysis, Rand Journal of Economics 25, 319–333.

——— , and Amit Seru, 2017, The use and misuse of patent data: Issues for corporate finance and beyond, NBER Working Paper 24053.

Lowry, Michelle, 2003, Why does IPO volume fluctuate so much?, Journal of Financial Economics 67, 3–40.

Maats, Frederiske, Andrew Metrick, Ayako Yasuda, Brian Hinkes, and Sofia Vershovski, 2011, On the consistency and reliability of venture capital databases, Working Paper.

Mokyr, Joel, 1990, Punctuated equilibria and technological progress, American Economic Review 80, 350–354.

Nanda, Ramana, and Matthew Rhodes-Kropf, 2013, Investment cycles and startup innovation, Journal of Financial Economics 110, 403–418.

Packalen, Mikko, and Jay Bhattacharya, 2018, New ideas in invention, NBER Working Paper 202922.

Pastor, Lubos, and Pietro Veronesi, 2005, Rational IPO waves, Journal of Finance pp. 1713–1757.

Phillips, Gordon, and Alexei Zhdanov, 2013, R&D and the incentives from merger and acquisition activity, Review of Financial Studies 34-78, 189–238.

Poulsen, Annette, and Mike Stegemoller, 2008, Moving from private to public ownership: Selling out to public firms versus initial public offerings, Financial Management pp. 81–101.

Ritter, Jay, 2019, Initial public offerings: Updated statistics, University of Florida.

Sebastiani, Fabrizio, 2002, Machine learning in automated text categorization, ACMCS 34, 1–47.

Trajtenberg, Manuel, Rebecca Henderson, and Adam Jaffe, 1997, University versus corporate patents: A window on the basicness of invention, Economics of Innovation and New Technology 5, 19–50.

Tushman, Michael L, and Philip Anderson, 1986, Technological discontinuities and organizational environments, Administrative Science Quarterly pp. 439–465.

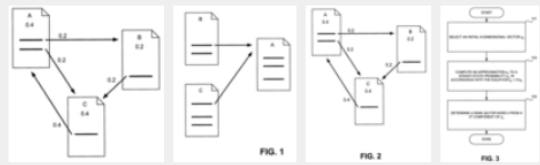
Figure I: Example of a Google Patent page
This figure shows the structure of a Google Patent page. The depicted patent is 6,285,999, commonly known as PageRank. Available at https://patents.google.com/patent/US6285999.

Method for node ranking in a linked database

Abstract

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

Images (4)



Classifications

G06F17/30864 Retrieval from the Internet, e.g. browsers by querying, e.g. search engines or meta-search engines, crawling techniques, push systems

G06F17/30728 Information retrieval; Database structures therefor ; File system structures therefor of unstructured textual data based on associated metadata or manual classification, e.g. bibliographic data using citations

Y10S707/99935 Query augmenting and refining, e.g. inexact access

Y10S707/99937 Sorting

Hide more classifications

US6285999B1
US Grant

Download PDF Find Prior Art

Inventor: Lawrence Page

Current Assignee: Leland Stanford Junior University , Google LLC

Original Assignee: Leland Stanford Junior University

Priority date: 1997-01-10

Family: US (10)

Date	App/Pub Number	Status
1998-01-09	US09004827	Expired - Lifetime
2001-09-04	US6285999B1	Grant
Show 8 more applications		
2012	US13616965	Expired - Lifetime

Info: Patent citations (28), Non-patent citations (20), Cited by (812), Legal events, Similar documents, Priority and Related Applications

External links: USPTO, USPTO Assignment, Espacenet, Global Dossier, Discuss

Description

CROSS-REFERENCES TO RELATED APPLICATIONS

This application claims priority from U.S. provisional patent application Ser. No. 60/035,205 filed Jan. 10, 1997, which is incorporated herein by reference.

STATEMENT REGARDING GOVERNMENT SUPPORT

This invention was supported in part by the National Science Foundation grant number IRI-9411306-4. The Government has certain rights in the invention.

FIELD OF THE INVENTION

This invention relates generally to techniques for analyzing linked databases. More particularly, it relates to methods for assigning ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database.

BACKGROUND OF THE INVENTION

Due to the developments in computer technology and its increase in

Claims (29)

What is claimed is:

1. A computer implemented method of scoring a plurality of linked documents, comprising:

obtaining a plurality of documents, at least some of the documents being linked documents, at least some of the documents being linking documents, and at least some of the documents being both linked documents and linking documents, each of the linked documents being pointed to by a link in one or more of the linking documents;

assigning a score to each of the linked documents based on scores of the one or more linking documents and

processing the linked documents according to their scores.

2. The method of claim **1**, wherein the assigning includes:

identifying a weighting factor for each of the linking documents, the weighting factor being dependent on the number of links to the one or more linking documents, and

Figure II: Time Series of Aggregate RETech

This figure reports the evolution of *RETech* for the aggregate patent corpus from 1930 to 2010. The aggregate stock reported below is the average RETech across all patents for the prior 20 quarters after applying a 5% quarterly rate of depreciation. All series are reported as four quarter moving averages.

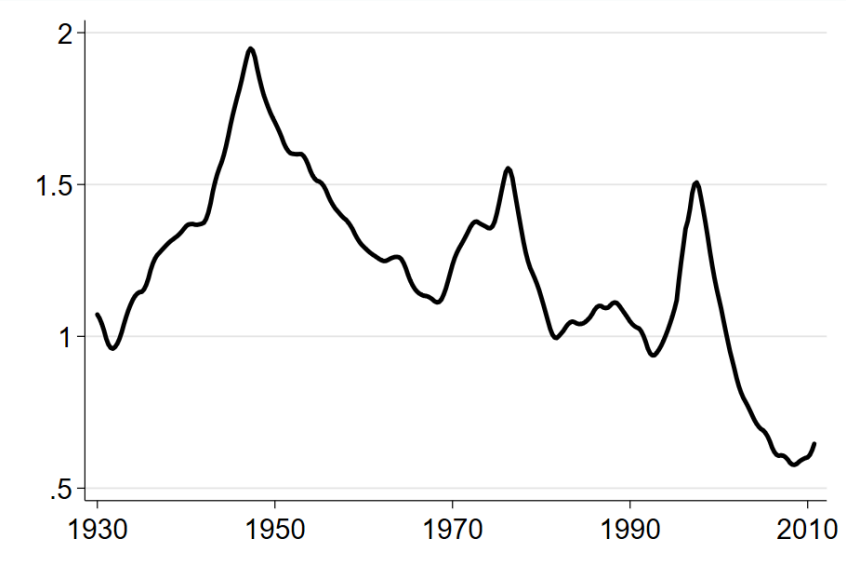


Figure III: Time Series of RETech, by Technology Category

This figure reports the evolution of *RETech* across six technology categories from 1930 to 2010. The stock for each category is the average RETech across all patents in that category for the prior 20 quarters after applying a 5% quarterly rate of depreciation. All series are reported as four quarter moving averages. Next to each header, we report in parentheses the fraction of patents over the entire sample belonging to that category.

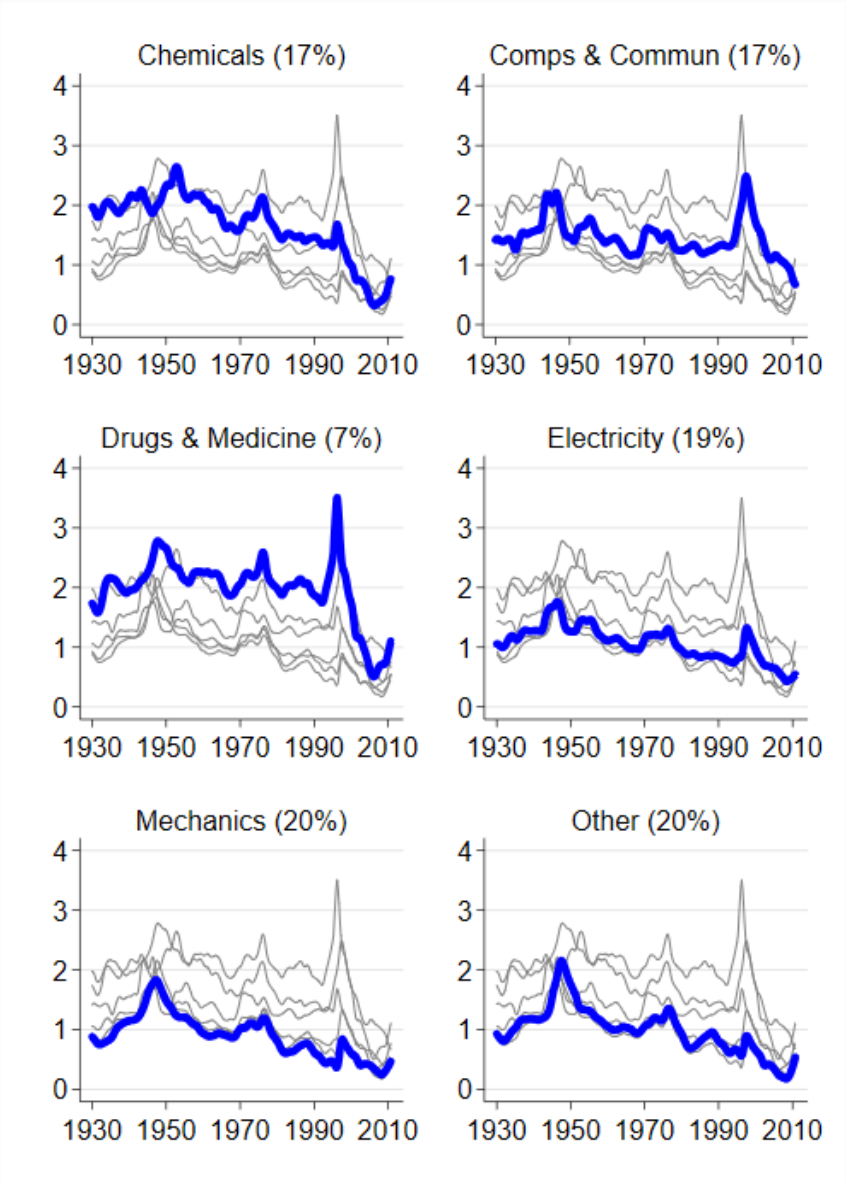


Figure IV: RETech, by Assignee Type

This figure reports the evolution of *RETech* across four categories of assignee from 1930 to 2010. The stock for each category is the average RETech across all patents in that category for the prior 20 quarters after applying a 5% quarterly rate of depreciation. All series are reported as four quarter moving averages.

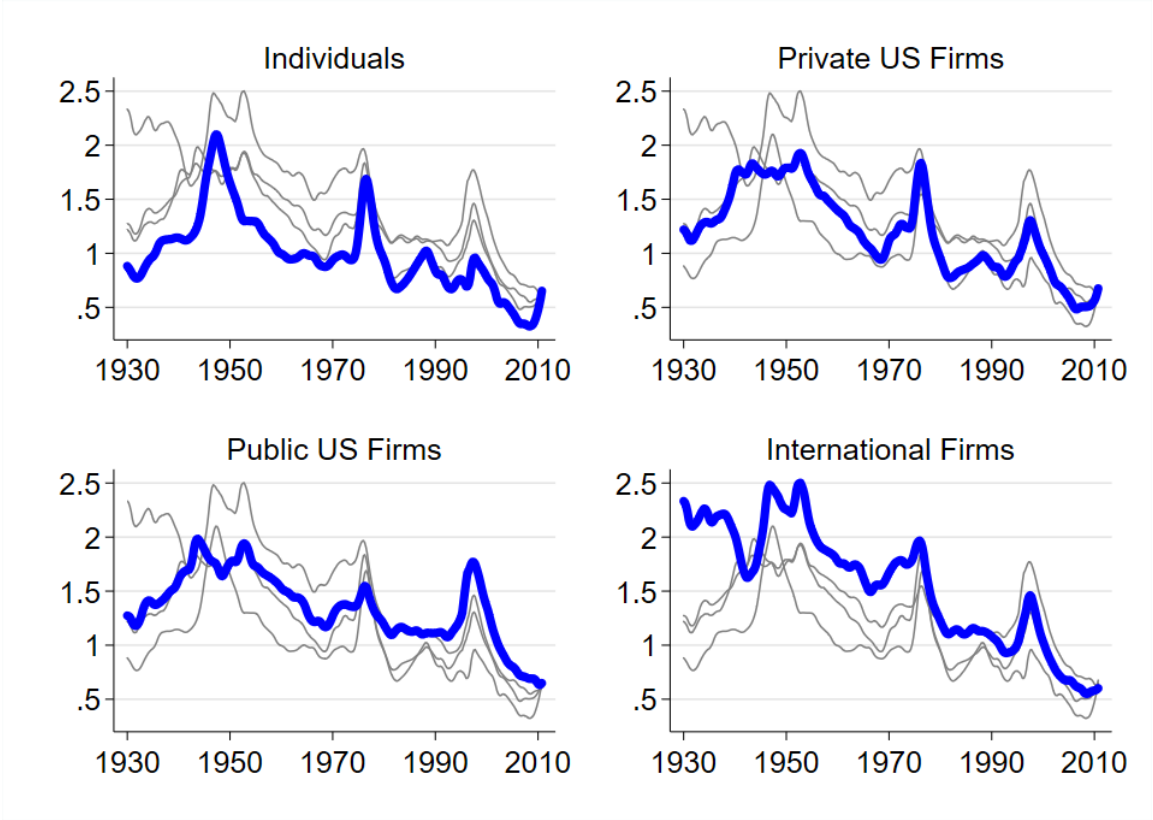
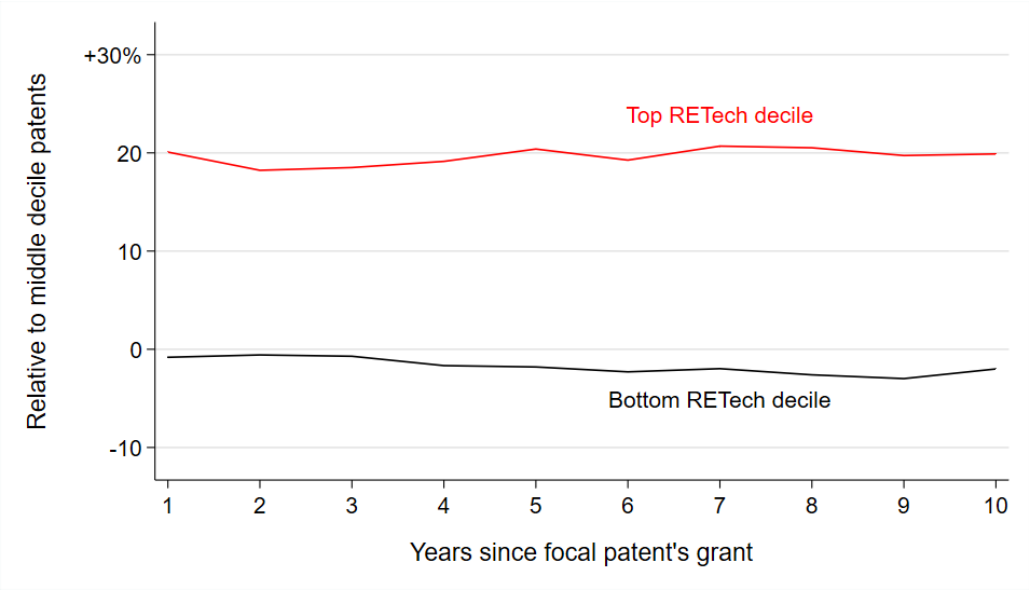


Figure V: High RETech patents as substitutes

This figure reports how the citations of "follow-on"patents differ for patents in the top and bottom deciles of *RETech* for patents from 1930 to 2000. A "follow-on" patent is a patent that cites a focal patent or any of its predecessors where a predecessor is a patent cited by the focal patent. Following Funk and Owen-Smith (2016), we compute, for each year following the focal patent's grant, the fraction of future patents that (Panel A) cite the focal patent but none of the patents it cited (indicating the focal patent substitutes for its predecessors), versus (Panel B) the fraction that cite the focal patent together with any of the patents it cited (indicating the focal patent complements its predecessors). We separately compute these fractions for focal patents in the top, bottom, and middle (5th and 6th) deciles of *RETech*, and report the percent difference from the middle decile to the top and bottom deciles.

Panel A: Indicating substitution



Panel B: Indicating complementarity

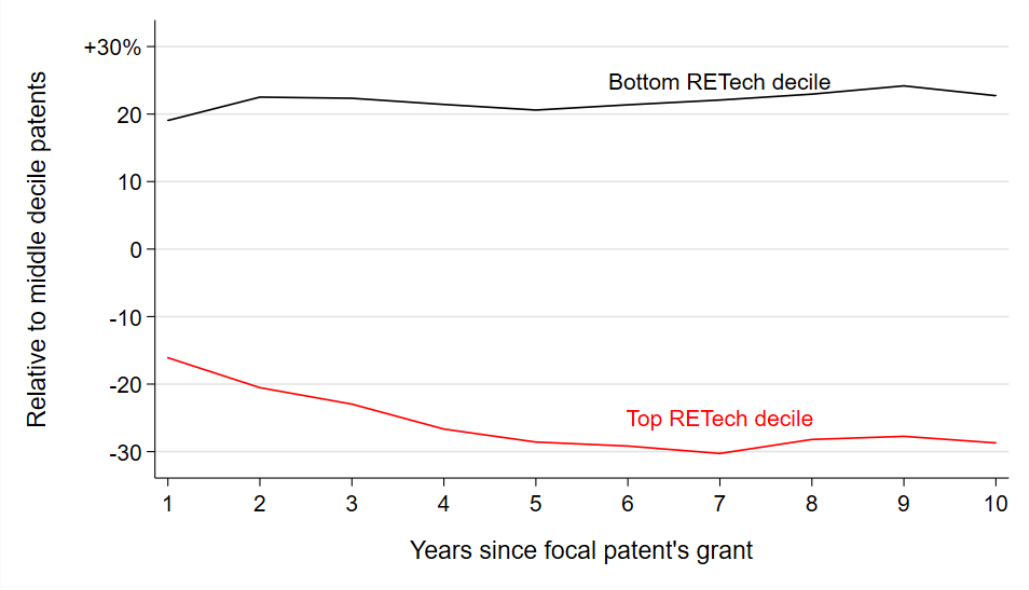


Figure VI: Distribution of RETech at exit
This figure shows the distribution of startup-level RETech in the periods when startups exit via IPO or sell-out.

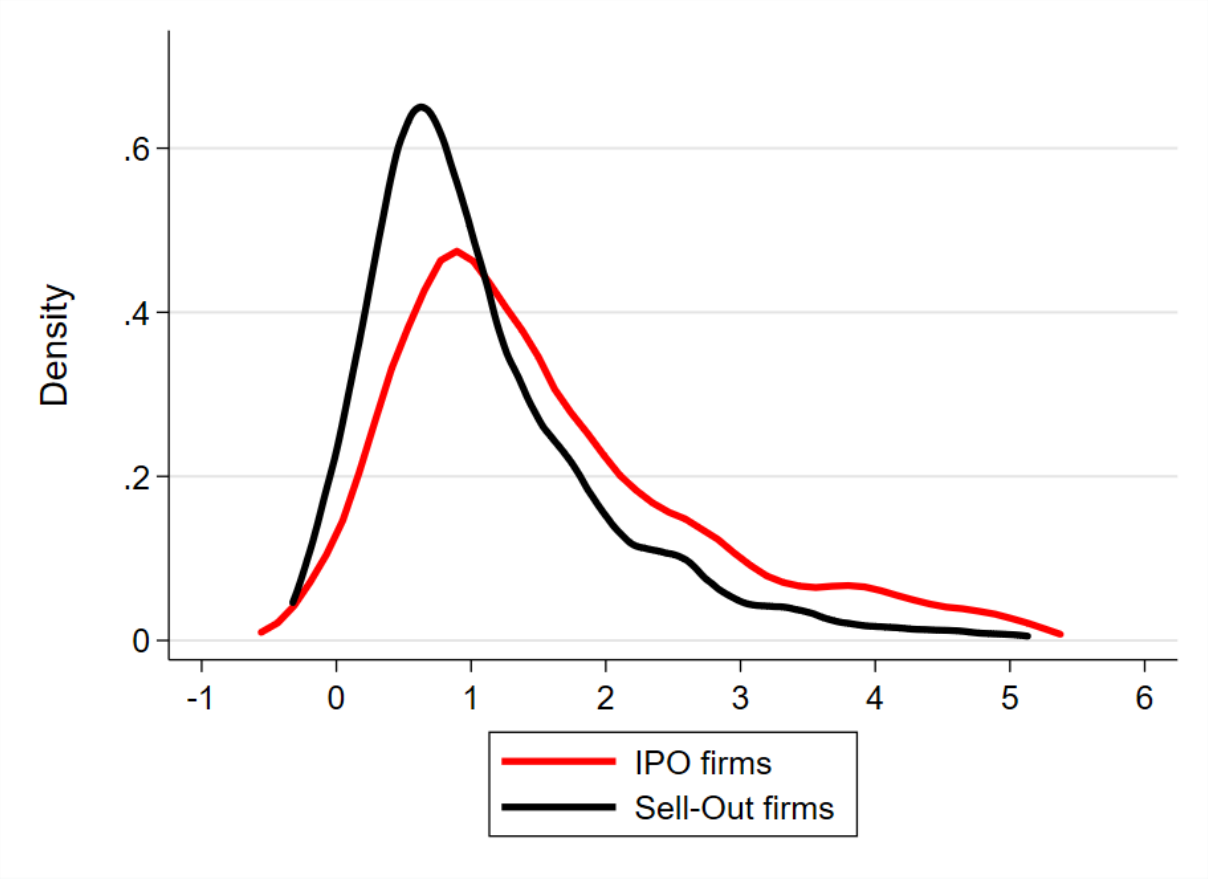


Figure VII: Fraction of exits that are IPOs, by RETech quintile
This figure shows the fraction of startups that exit via IPO (as opposed to sell-outs) as a function of their pre-exit RETech. We sort exiting startups that have applied for at least one patent in the proceeding five years into quintiles of their stock of RETech as of the quarter prior to exit.

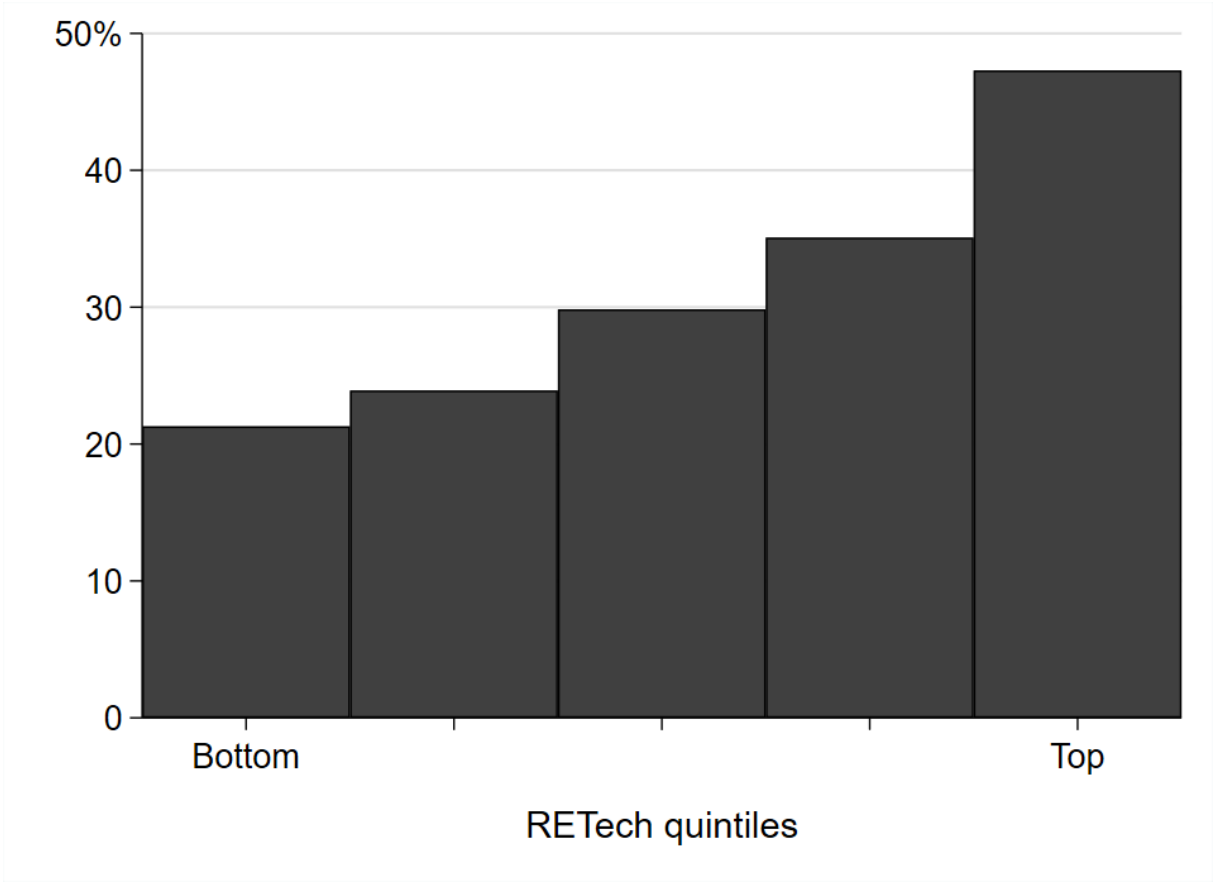


Figure VIII: Time since first VC round, as of exit

This figure shows the average time (in quarters) between a startup's first VC round and its exit via IPO or sell-out across quintiles of RETech. Quintiles are based on the average RETech of patents applied for before the startup exits. Error bars report the standard error of the average exit speed for a given quintile and exit.

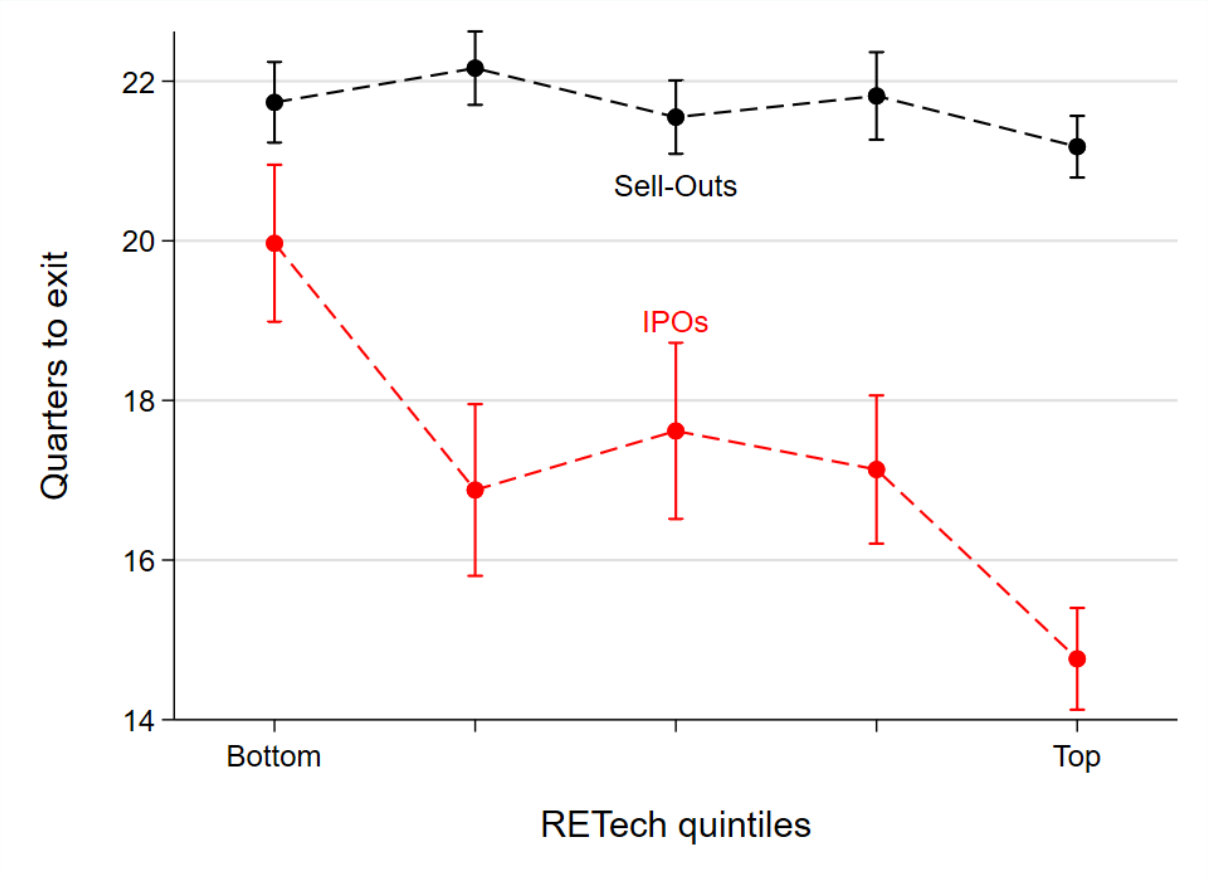


Figure IX: Startup-Acquirer Matching

This figure reports the number of sell-outs taking place between a startup and a public acquirer as a function of the RETech of each entity as of the date of the deal. We compute the RETech of public acquirers by aggregating their patents following the same aggregation procedure we use for startups. We then divide both selling startups and acquirers into 25 RETech bins and plot the number of deals in each of the 625 pairs of bins. The number of deals in a given bin is encoded by both size and color. Empty bins are not plotted. The overlaid line is the predicted *Acquirer RETech* from regressing it on *Target RETech*.

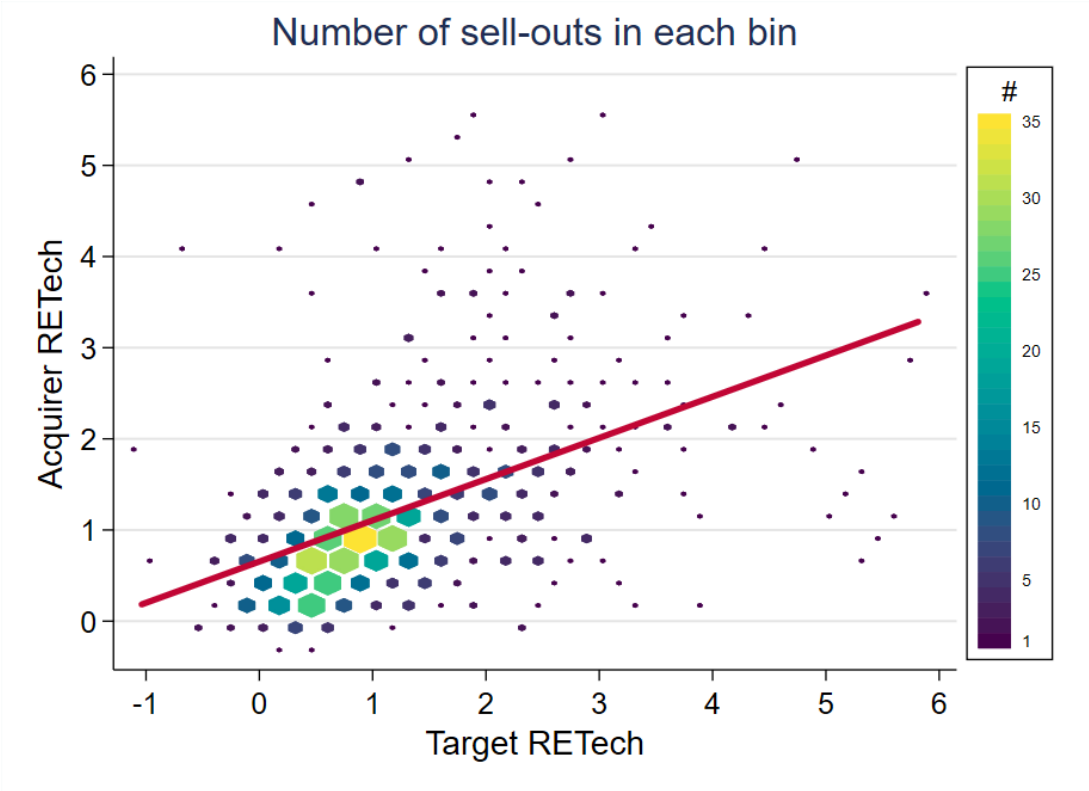


Table I: Changes in Patent Word Usage: Examples

This table reports, for five years between 1930 and 2010, lists the ten words that have the largest year-over-year increase (Panel A) or decrease (Panel B) (Δ as in Section II.B) in use across all patents. For the illustrative purpose of this table, for each year, we remove "common" words (those used in more than 25% of patents), new or disappearing words, and "rare" words (those used in fewer than 100 patents across the year and the prior year). After these filters, to focus on words that are prevalent enough for readers to recognize, we keep the top decile of words based on the number of patents the word is used in across the year and the prior year.

Panel A: Words with largest increase in use

1935	**1975**	**1985**	**1995**	**2005**
cent	bolts	laser	polypeptides	broadband
leaves	effort	japanese	deletion	intervening
axes	lithium	wavelength	clones	candidates
packing	user	publication	polypeptide	click
column	describes	blood	peptides	configurable
lead	exemplary	infrared	recombinant	luminance
coupled	entitled	polymer	cdna	abstract
notch	typically	mount	nucleic	acquiring
copper	phantom	optical	transcription	telecommunications
chain	exploded	comparative	plasmid	gamma

Panel B: Words with largest decline in use

1935	**1975**	**1985**	**1995**	**2005**
chambers	assistant	sulfuric	cassette	vegetable
crank	inventor	collection	ultrasonic	acyl
boiling	inventors	crude	machining	spiral
agent	firm	stock	abutment	gram
seats	priority	dioxide	tape	wedge
yield	john	evident	sand	gelatin
reducing	foreign	hydrocarbon	packing	crude
engine	sept	shut	bottle	oven
bell	june	circuitry	slidable	maybe
film	corporation	oxides	insofar	drilling

Table II: Summary Statistics: Patent-level Sample

This table presents summary statistics for patent applications between 1930 and 2010. All variables are defined in Table A1. P25 and P75 denote the 25th and 75th percentiles.

	N	Mean	SD	P25	Median	P75
RETech	6,594,248	1.64	1.81	0.51	1.27	2.34
Tech Breadth	6,594,143	0.42	0.22	0.24	0.47	0.60
Private Similarity	6,594,248	0.15	0.05	0.12	0.15	0.18
LI Similarity	6,594,248	0.11	0.05	0.06	0.09	0.13
Foreign Similarity	6,594,248	0.15	0.06	0.11	0.14	0.19
Originality	5,335,987	0.40	0.33	0.00	0.46	0.67
# of Cites	6,595,226	1.58	2.91	0.00	1.00	2.00
Scope	4,541,529	4.14	2.82	2.00	3.00	5.00
NumClaims	4,541,456	14.96	10.98	7.00	13.00	20.00
KPSS Value	1,781,386	9.75	23.69	0.73	3.25	9.16
Subs	5,548,289	0.24	0.28	0.00	0.13	0.37
Comp	5,548,289	0.03	0.08	0.00	0.00	0.03
RETech(New)	6,594,248	0.31	0.71	0.00	0.00	0.35
RETech(Established)	6,594,248	1.32	1.45	0.41	1.10	2.00

Table III: Summary Statistics: Patent-level Sample

This table presents correlations between patent-level variables for patent applications between 1930 and 2010. All variables are defined in Table A1. All pairwise correlations are statistically significant at the 1% level.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	RETech	1.00											
(2)	Tech Breadth	-0.32	1.00										
(3)	Private Similarity	-0.16	0.32	1.00									
(4)	LI Similarity	0.12	-0.14	0.41	1.00								
(5)	Foreign Similarity	-0.04	0.14	0.62	0.74	1.00							
(6)	Originality	-0.00	0.04	0.06	0.04	0.02	1.00						
(7)	Log(1+# of Cites)	0.13	-0.09	0.02	0.14	0.02	-0.00	1.00					
(8)	Log(1+KPSS Value)	0.08	-0.03	-0.01	-0.07	-0.22	0.02	0.08	1.00				
(9)	Subs	0.20	-0.06	0.01	0.12	0.05	-0.05	0.39	0.03	1.00			
(10)	Comp	-0.04	0.09	0.03	-0.08	-0.03	-0.06	0.14	0.03	-0.05	1.00		
(11)	Scope	0.04	-0.04	-0.03	0.07	0.09	0.08	0.08	0.01	0.03	-0.02	1.00	
(12)	NumClaims	0.02	-0.09	-0.00	0.03	-0.04	0.04	0.18	0.11	-0.01	-0.01	0.06	1.00

Table IV: Discontinuities in Technological Evolution

This table presents regressions on a sample of patents applications between 1930 and 2000. Technology substitutes (*Subs*) is the percentage of follow-on patents that do not cite any of the focal patent's predecessors. Technology complements (*Comp*) is the percentage of follow-on patents that cite the focal patent and at least one of the focal patent's predecessors. A follow-on patent is a patent that cites a focal patent or any of the patents cited therein, which we call "predecessors". To facilitate interpretation, RETech is standardized. Fixed effects are included based on a patent's grant year (cohort) and NBER-technology category (technology). Adjusted R^2 is reported as a percentage. Standard errors are clustered by the patent's grant year and t-stats are reported in parentheses. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Subs (1)	Comp (2)	Subs (3)	Comp (4)
RETech	1.804*** (9.88)	-0.263*** (-16.72)	1.562*** (8.28)	-0.234*** (-14.51)
Cohort FE	Yes	Yes		
Technology FE	Yes	Yes		
Cohort X Technology FE			Yes	Yes
Assignee Type FE	Yes	Yes	Yes	Yes
Observations	3,841,530	3,841,530	3,841,488	3,841,488
R2 (%)	3.9	2.3	4.7	2.5

Table V: Breakthrough Patents

This table reports summary statistics for the percentiles of a comprehensive list of patents identified by Kelly, Papanikolaou, Seru, and Taddy (2019) (henceforth, KPST) based on several on-line lists of "important" patents. We focus on the subset of patents applied for between 1930 and 2010 over which the textual measures in this paper are defined. These patents are listed in detail in Table A3. Percentiles are cohort-adjusted, i.e., we remove year fixed effects before computing percentiles across all patents. The percentiles for the KPST measure are taken directly from their Table A.6. All variables are defined in Table A1. "Brdth" and "Orig" are short for *Tech Breadth* and *Originality*, respectively.

Patent	RETech	Cites	KPSS	KPST	Brdth	Orig
Average:	0.72	0.75	0.69	0.83	0.52	0.55
Median:	0.82	0.88	0.80	0.89	0.53	0.62
Std error:	(0.03)	(0.03)	(0.05)	(0.02)	(0.03)	(0.04)
N:	101	101	30	101	101	62
Ex ante measureable:	Yes	No	Yes	No	Yes	Yes

Table VI: Summary Statistics: Startup Sample

This table presents summary statistics for the quarterly sample of venture-backed startups between 1980 and 2010. Startups are in the sample from their founding date until the quarter of their final outcome. All variables are defined in Table A1. P25 and P75 denote the 25th and 75th percentiles.

	N	Mean	SD	P25	Median	P75
RETech	347,929	0.66	1.14	0.00	0.00	0.98
Tech Breadth	347,929	0.13	0.18	0.00	0.01	0.27
Private Similarity	347,929	0.06	0.06	0.00	0.06	0.11
LI Similarity	347,929	0.04	0.05	0.00	0.03	0.07
Foreign Similarity	347,929	0.05	0.06	0.00	0.04	0.09
Log(1+Firm Age)	347,929	3.07	1.15	2.40	3.18	3.76
No PatApps	347,929	0.47	0.50	0.00	0.00	1.00
Log(1+PatApps)	347,929	0.79	0.97	0.00	0.69	1.39
Log(MTB)	347,929	0.15	0.08	0.11	0.15	0.19
MKT Return	347,929	0.01	0.13	-0.08	0.02	0.09
Log(1+Cites)	347,929	0.54	0.70	0.00	0.00	1.02
Q4	347,929	0.25	0.43	0.00	0.00	0.00
Originality	347,929	0.16	0.20	0.00	0.00	0.31
log(CumVCFunding)	347,929	5.40	4.66	0.00	7.55	9.62
No Funding	347,929	0.37	0.48	0.00	0.00	1.00
VC market share	347,929	0.00	0.01	0.00	0.00	0.00
Scope	347,929	1.55	1.95	0.00	0.93	2.62
NumClaims	347,929	8.08	10.34	0.00	3.99	13.85
IPO rate (x100)	347,929	0.42	6.43	0.00	0.00	0.00
Sell-Out rate (x100)	347,929	0.73	8.50	0.00	0.00	0.00
Failure Rate (x100)	347,929	0.33	5.75	0.00	0.00	0.00
Remain Rate (x100)	347,929	98.53	12.05	100.00	100.00	100.00

Table VII: The Determinants of Startups' Exits - Baseline

This table presents cross-sectional tests relating startups' ex ante technological traits to their IPO and sell-out exits. The sample is a quarterly panel of VC-backed startups from 1980-2010. Competing risk hazard, logit, and multinomial models report exponentiated coefficients minus one. OLS models report coefficients scaled by 100. In column (5) of Panel A, we restrict the sample to observations where a startup is going public or selling-out. Independent variables are lagged one quarter and non-binary controls are standardized for convenience. All variables are defined in Table A1. *LI Similarity* and *Foreign Similarity* are orthogonalized relative to *Private Similarity*. Industry and location fixed effects are based on the major industry group and state, respectively, reported in VentureXpert. Technology fixed effects are based on the most common NBER-technology category across a firm's patents. Adjusted R^2 is reported as a percentage. Standard errors are clustered by startup (except for column (5) in Panel A, which is clustered by year) and t-stats are reported in parentheses. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Baseline Models

	Competing Risk Hazard		OLS (coefficients x100)		
	IPO	Sell-Out	IPO	Sell-Out	IPO vs Sell-Out
	(1)	(2)	(3)	(4)	(5)
RETech	0.247***	-0.143***	0.073***	-0.059***	2.176*
	(12.26)	(-5.53)	(3.60)	(-3.07)	(2.03)
Tech Breadth	0.539***	-0.092**	0.095***	-0.109***	3.590**
	(10.30)	(-2.30)	(3.39)	(-3.45)	(2.30)
Private Similarity	0.141*	-0.384***	0.031	-0.387***	4.975**
	(1.83)	(-6.72)	(0.83)	(-8.46)	(2.48)
LI Similarity	0.233***	-0.033	0.038	-0.095***	1.188
	(4.26)	(-0.76)	(1.23)	(-2.69)	(0.85)
Foreign Similarity	-0.064	0.027	-0.011	0.075***	-1.333
	(-1.59)	(0.88)	(-0.60)	(3.09)	(-1.15)
No PatApps	1.895***	-1.918***	0.400***	-1.705***	40.736***
	(12.15)	(-19.04)	(5.79)	(-15.31)	(7.07)
Log(1+PatApps)	0.267***	-0.162***	0.124***	-0.176***	6.859***
	(7.49)	(-5.50)	(5.19)	(-5.92)	(4.79)
Log(MTB)	0.097***	0.061***	0.164***	0.063	3.830
	(3.42)	(2.72)	(4.56)	(1.17)	(1.28)
MKT Return	0.339***	0.023	0.045***	0.035*	1.942**
	(12.49)	(1.06)	(3.46)	(1.71)	(2.05)
Log(1+Cites)	0.148***	0.117***	0.058***	0.151***	-0.612
	(3.53)	(4.01)	(2.77)	(5.25)	(-0.60)
Q4	-0.106	0.053	0.061	0.258***	-2.890
	(-1.47)	(1.02)	(1.28)	(4.03)	(-1.33)
Originality	-0.121***	-0.143***	-0.023	-0.136***	1.497
	(-2.90)	(-4.71)	(-1.18)	(-5.64)	(1.48)
log(CumVCFunding)	0.326***	0.656***	0.258***	0.534***	3.980***
	(4.27)	(10.30)	(8.24)	(14.30)	(2.80)
No Funding	-0.899***	-2.490***	-0.123**	0.139**	36.561***
	(-5.14)	(-9.70)	(-2.18)	(2.13)	(5.38)
VC market share	0.072***	-0.008	0.044**	0.045**	-1.442
	(4.75)	(-0.54)	(2.34)	(2.45)	(-1.56)

Continued on next page

	Competing Risk Hazard		OLS		
	IPO	Sell-Out	IPO	Sell-Out	IPO vs Sell-Out
	(1)	(2)	(3)	(4)	(5)
Scope	0.104***	-0.154***	0.021	-0.104***	2.823**
	(2.86)	(-4.32)	(1.00)	(-4.39)	(2.22)
NumClaims	-0.121***	-0.154***	-0.014	-0.122***	1.817*
	(-2.83)	(-4.56)	(-0.80)	(-5.12)	(1.92)
Year FE	No	No	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes	Yes
Technology FE	No	No	Yes	Yes	Yes
Location FE	No	No	Yes	Yes	Yes
Firm Age FE	No	No	Yes	Yes	Yes
Firm Cohort FE	No	No	Yes	Yes	Yes
Observations	346,490	345,403	342,146	342,146	3,913
R2 (%)	N/A	N/A	0.7	0.9	31.5

Panel B: Logit and Multinomial Models

	Logit		Multinomial Logit	
	IPO	Sell-Out	IPO	Sell-Out
	(1)	(2)	(3)	(4)
RETech	0.263***	-0.070***	0.262***	-0.069***
	(12.14)	(-3.21)	(12.08)	(-3.11)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	No	No	No	No
Observations	342,157	342,157	342,157	
R2 (%)	6.8	10.2	N/A	

Table VIII: The Determinants of Startups' Exits - Dynamic Responses

This table presents dynamic cross-sectional tests relating startups' ex ante technological traits to their exit over several horizons. In Panel A, column 1 repeats the OLS model examining IPO exits from column (3) in Table VII. The remaining columns replace the one-period ahead IPO exit indicator with longer horizons. We repeat this analysis for sell-outs in Panel B and remaining private in Panel C. In all models, the sample, independent variables, and coefficient interpretation are the same as the OLS models in Table VII. Coefficients are scaled by 100 and should be interpreted as additive percentage points shifts in the outcome for a one standard deviation increaase in RETech. For brevity, the control variables and fixed effects are omitted. Standard errors are clustered by startup and t-stats are reported in parentheses. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Horizon:	Qtr	Year	2 Years	3 Years	4 Years	5 Years
			Panel A: Exit by IPO within horizon			
RETech	0.073***	0.264***	0.360***	0.344**	0.368*	0.344
	(3.60)	(3.44)	(2.70)	(2.01)	(1.82)	(1.53)
R2 (%)	0.7	2.7	4.7	6.2	7.2	8.0
			Panel B: Exit by Sell-Out within horizon			
RETech	-0.059***	-0.166**	-0.114	-0.051	-0.026	-0.001
	(-3.07)	(-2.18)	(-0.74)	(-0.23)	(-0.10)	(-0.00)
R2 (%)	0.9	3.5	6.5	8.7	10.1	10.8
			Panel C: Still Private at end of horizon			
RETech	-0.043	-0.219*	-0.460**	-0.568**	-0.644**	-0.648*
	(-1.44)	(-1.93)	(-2.24)	(-2.12)	(-2.08)	(-1.95)
R2 (%)	1.4	5.7	10.7	14.9	18.4	21.4
Observations	342,146	342,146	342,146	342,146	342,146	342,146

Table IX: Robustness of Baseline Results

This table presents robustness tests of the main results in Table VII. For brevity, we only report the main coefficient on *RETech* for each test. Each row (Test #) corresponds to an alteration of the main test. Except for the listed alteration, each of the models within a row repeats the corresponding model in the same column of Table VII. To facilitate interpretation, competing risk models report exponentiated coefficients minus one, OLS models report coefficients scaled by 100, and *RETech* is standardized and lagged one quarter. Standard errors are clustered by startup unless otherwise noted and t-stats are reported in parentheses. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

		Competing Risk Hazard		OLS	
Test #	Test alteration	IPO (1)	Sell-Out (2)	IPO (3)	Sell-Out (4)
(1)	Include: VC fixed effects (Lead VC of previous round)			0.062*** (2.97)	-0.056*** (-2.87)
(2)	Year X Industry fixed effects			0.073*** (3.63)	-0.054*** (-2.77)
(3)	Cluster SE by startup technology	0.247*** (8.92)	-0.143*** (-4.39)	0.073*** (6.65)	-0.059*** (-3.56)
(4)	Cluster SE by year	0.247*** (7.75)	-0.143*** (-4.70)	0.073*** (2.83)	-0.059** (-2.21)
(5)	Cluster SE by startup cohort	0.247*** (7.88)	-0.143*** (-4.31)	0.073*** (3.01)	-0.059*** (-2.68)
(6)	Only controls: *No PatApps* and *Log(1+PatApps)*	0.258*** (13.04)	-0.122*** (-4.90)	0.080*** (3.97)	-0.040** (-2.12)
(7)	Exclude *Log(1+Cites)* as control	0.267*** (15.03)	-0.098*** (-4.70)	0.091*** (4.70)	-0.041** (-2.27)
(8)	Include: Controls for bank financing and patent generality	0.229*** (10.85)	-0.138*** (-5.28)	0.071*** (3.53)	-0.054*** (-2.77)
(9)	Code sell-outs <$25m (09$) as failures	0.247*** (12.26)	-0.134*** (-4.72)	0.073*** (3.60)	-0.045*** (-2.43)
(10)	Code sell-outs when buyer is financial acquirer as failure	0.247*** (12.26)	-0.141*** (-5.38)	0.073*** (3.60)	-0.060*** (-3.19)
(11)	Observations in [1980, 1995] N=112,634; Firms=3,951	0.134*** (4.31)	-0.328*** (-5.00)	0.063 (1.38)	-0.154*** (-5.12)
(12)	Observations in [1996, 2010] N=229,449, Firms=7,483	0.170*** (4.43)	-0.066** (-2.26)	0.060*** (2.90)	-0.013 (-0.50)
(13)	Sample and controls constructed annually	0.229*** (11.02)	-0.066*** (-2.86)	0.214*** (2.66)	-0.226*** (-2.89)

Table X: The Determinants of Startups' Exits - Decomposition

This table presents cross-sectional tests relating startups' ex ante technological traits to their exit. Each of the models repeats the corresponding model from Table VII, but replaces the main variable *RETech* with a decomposition focusing on two subsets of words in the patent corpus for each year. *RETech(Established)* is defined in Equation 6 and only includes words that are at least ten years old in a given year. *RETech(New)* is defined in Equation 7 and only includes words that are less than ten years old in a given year. For brevity, we only report the coefficients on the decomposed variables. To facilitate interpretation, competing risk models report exponentiated coefficients minus one, OLS models report coefficients scaled by 100, and *RETech(Established)* and *RETech(New)* is standardized and lagged one quarter. Independent variables are lagged one quarter. Adjusted R^2 is reported as a percentage. Standard errors are clustered by startup and t-stats are reported in parentheses. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Competing Risk Hazard		OLS	
	IPO	Sell-Out	IPO	Sell-Out
	(1)	(2)	(3)	(4)
RETech(Established)	0.312***	-0.100***	0.110***	-0.037*
	(10.37)	(-4.00)	(5.28)	(-1.75)
RETech(New)	0.012	-0.062**	-0.019	-0.029
	(0.51)	(-2.54)	(-1.19)	(-1.59)
Controls	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes
Industry FE	No	No	Yes	Yes
Technology FE	No	No	Yes	Yes
Location FE	No	No	Yes	Yes
Firm Age FE	No	No	Yes	Yes
Firm Cohort FE	No	No	Yes	Yes
Observations	346,490	345,403	342,146	342,146
R2 (%)	N/A	N/A	0.7	0.9

Table XI: Ancillary Evidence: Acquirer CARs and IPO shares

This table presents tests based on subsamples of VC-backed startups that are acquired by publicly traded firms or go public and have applied for patents within five years before the acquisition announcement or public listing. Acquirer CAR is the cumulative abnormal returns from a Fama-French 3 factor model estimated over 100 days (70+ are required), with a 50 day gap before the announcement. *Primary Share %* is the percentage (between 0 and 100) of IPO shares that are newly issued (as opposed to existing shares sold by insiders) and based on data from SDC. *Only Primary Shares* equals one when all shares are newly issued and zero if SDC records any sales by insiders. Adjusted R^2 is reported as a percentage. *RETech* is measured as of the quarter of the merger announcement or initial listing. We include year fixed effects for the year of exit. Standard errors are clustered by target industry as defined by SDC in the sell-out sample and by two digit SIC in the IPO sample. t-stats are reported in parentheses. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Sample:	Sell-out to public firm		IPO	
Dep. Variable:	Acquirer CAR[-1,1]	Acquirer CAR[-3,3]	%Primary Shares	Only Primary Shares
	(1)	(2)	(3)	(4)
RETech	-0.005*	-0.006**	1.513***	0.042***
	(-1.83)	(-2.15)	(4.78)	(4.72)
Year FE	Yes	Yes	Yes	Yes
Observations	905	905	1,167	1,167
R2 (%)	1.4	1.6	8.4	9.1

Table A1: Variable Definitions

Patent-Level Variables

RETech	See Equation 3 and Section II.B.
Tech Breadth	See Equation 4 and Section II.C.
LI Similarity	See Equation 5 and Section II.C.
Private Similarity	Similar to *LI Similarity*. See Section II.C.
Foreign Similarity	Similar to *LI Similarity*. See Section II.C.
KPSS Value	From Kogan, Papanikolaou, Seru, and Stoffman (2016).
# of Cites	Number of citations received in the first five years after publication by the USPTO. Citations up to December 31, 2013.
Originality	The originality of a focal patent is defined as 1 minus the HHI of the technology fields of the patents cited by the focal patent (Trajtenberg, Henderson, and Jaffe (1997)). We use the adjustment given in Hall, Jaffe, and Trajtenberg (2001) to reduce bias for patents that contain few backward citations. We convert U.S. Patent Classifications to the NBER technology codes so that *Tech Breadth* and *Originality* are based on the same granularity of technology classifications.
RETech (Established)	See Equation 6 and Section IV.F.
RETech (New)	See Equation 7 and Section IV.F.
Subs	The percentage of follow-on patents that do not cite the patent. Follow-on patents are any patents that cite the patent itself or any of the patents it cites. The application date of follow-on patents must be within the 10 years following the patent.
Comps	The percentage of follow-on patents that cite the focal patent and at least one of the focal patents predecessors. The application date of follow-on patents must be within the 10 years following the patent.
Scope	Number of USPC subclasses in *uspc_current.tsv* (PatentsView.org).
NumClaims	Number of claims from *patent.tsv* (PatentsView.org).

Startup-Quarter Variables

RETech	The average RETech of patents from the last 20 quarters after applying a 5% quarterly rate of depreciation. Formally, for firm i in quarter q, $RETech_(i,q) = \frac{1}{\lvert P_(i,q)\rvert} \sum_{p \in P_(i,q)} (1-d)^{k(p)} RETech(p)$ where $P_(i,q)$ is the set of patents firm i applied for in [q-19,q], $k(p)$ is the number of quarters since patent p was applied for, and $RETech(p)$ is the RETech of patent p. Quarterly depreciation d is 5%.
Tech Breadth	Converted to startup-quarter from patent-level like *RETech*.
Private Similarity	Converted to startup-quarter from patent-level like *RETech*.
LI Similarity	Converted to startup-quarter from patent-level like *RETech*.
Foreign Similarity	Converted to startup-quarter from patent-level like *RETech*.
RETech (Established)	Converted to startup-quarter from patent-level like *RETech*.
RETech (New)	Converted to startup-quarter from patent-level like *RETech*.
Scope	Converted to startup-quarter from patent-level like *RETech*.
NumClaims	Converted to startup-quarter from patent-level like *RETech*.
Originality	Converted to startup-quarter from patent-level like *RETech*.
CiteStock	Converted to startup-quarter from patent-level *# of Cites* like *RETech*.
Log(1+Cites)	Log(1+CiteStock)
No PatApps	Dummy variable equal to one if the startup has not applied for a patent (which was eventually granted) during the prior 20 quarters.

No Funding	Dummy variable equal to one if the startup has not received a VC funding round during the prior 20 quarters.
Log(1+PatApps)	The # of (granted) patent applications in the prior 20 quarters.
IPO	One if the startup goes public in the quarter, zero before.
Sell-out	One if the startup is acquired in the quarter, zero before.
log(CumVCFunding)	Cumulutive VC funding received as of quarter q. Source: VentureXpert.
VC market share	The market share of VCs backing a startup's most recent round. VC-by-round funding data from SDC. Equals zero before first round.

Quarterly variables

Log(MTB)	Aggregate market-to-book is computed quarterly using all firms in the CRSP-Compustat database. We sum each subcomponent of MTB across all firms, then compute $MTB = (at - ceq + mve - txdb)/at$ as defined in Kaplan and Zingales (1997).
MKT Return	In quarter q, the market return for $q-2$ to $q-1$, inclusive, using geometric compounding of daily market returns from Ken French's daily factor file.
Q4	One if q is the fourth quarter, else zero.

Table A2: Years Between Keys Events for Ventured-Backed Startups

This table presents information on key events for startups in the main analysis sample described in Panel B of Table II and Section II.B. A startup's first patent is based on the earliest application date for (eventually) granted patents. Information on VC funding, timing, and exits are from VentureXpert, and patenting information is from Google Patents.

Panel A: Events after the startup's founding

		Years between the startup's founding and event				
Event	N (startups)	Mean	SD	P25	Median	P75
First patent	9,167	4.42	10.76	0.75	2.25	5.75
VC funding	9,167	5.29	10.63	0.50	1.75	5.50
IPO	1,677	9.41	9.89	4.50	7.00	11.25
Acquisition	3,377	11.23	10.50	6.00	8.50	12.75

Panel B: Events after the startup's first patent

		Years between the startup's first patent and event				
Event	N (startup)	Mean	SD	P25	Median	P75
VC funding	9,167	0.87	7.78	-2.00	-0.25	2.50
IPO	1,677	3.10	8.45	-0.50	3.00	6.75
Acquisition	3,377	7.46	7.00	3.75	6.25	10.00

Table A3: Percentiles of Various Statistics for a Sample of Important Patents

The patents below are the subset of key important patents listed in Kelly, Papanikolaou, Seru, and Taddy (2019) (henceforth, KPST) applied for between 1930 and 2010 over which the textual measures in this paper are defined. The percentiles for the KPST measure are taken directly from their Table A.6. The variables are defined in Table A1. "Brdth" and "Orig" are short for *Tech Breadth* and *Originality*, respectively. Percentiles are cohort-adjusted, i.e., we remove year fixed effects before computing percentiles across all patents.

Ex ante measureable:		Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Patent	Grant Year	RETech	Cites	KPSS	KPST	Brdth	Orig	Priv Simm	LI Simm	Frgn Simm
Panel A: Summary statistics of percentiles in Panel B										
Average:		0.72	0.75	0.69	0.83	0.52	0.55	0.51	0.62	0.56
Median:		0.82	0.88	0.80	0.89	0.53	0.62	0.46	0.68	0.57
Std error:		(0.03)	(0.03)	(0.05)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
N:		101	101	30	101	101	62	101	101	101
Panel B: Percentiles of various measures for breakthrough patents										
1,929,453	1933	0.95	0.66		0.98	0.26		0.82	0.70	0.90
1,941,066	1933	0.78	0.68		0.93	0.74		0.19	0.97	0.86
1,948,384	1934	0.78	0.66		0.87	0.79		0.21	0.76	0.66
1,949,446	1934	0.50	0.66		0.55	0.31		0.83	0.57	0.48
1,980,972	1934	1.00	0.65		0.98	0.18		0.84	0.79	1.00
2,021,907	1935	0.60	0.67		0.89	0.99		0.28	0.82	0.65
2,059,884	1936	0.94	0.66	0.63	0.59	0.46		0.41	0.72	0.89
2,071,250	1937	0.98	0.67	0.84	0.89	0.30		0.97	0.91	1.00
2,087,683	1937	0.86	0.66		0.92	0.68		0.37	0.84	0.91
2,153,729	1939	1.00	0.65		0.96	0.16		0.67	0.73	1.00
2,188,396	1940	0.78	0.63	0.60	1.00	0.17		0.83	0.62	0.80
2,206,634	1940	0.93	0.65		0.98	0.57		0.68	0.71	0.97
2,230,654	1941	0.93	0.62		0.93	0.25		0.88	0.75	0.51
2,258,841	1941	0.38	0.56		0.23	0.88		0.72	0.52	0.83
2,292,387	1942	0.52	0.56		0.95	0.94		0.58	0.81	0.63
2,297,691	1942	0.14	0.62		0.62	0.87		0.67	0.69	0.76
2,329,074	1943	0.91	0.89		0.56	0.20		0.99	0.95	0.98
2,390,636	1945	0.18	0.95		0.79	0.57		0.08	0.14	0.33
2,404,334	1946	0.59	0.97		0.23	0.80		0.83	0.79	0.96
2,436,265	1948	0.37	0.95		0.74	0.39	0.26	0.76	0.75	0.66
2,451,804	1948	0.92	0.93		0.74	0.20	0.06	0.94	0.80	0.50
2,495,429	1950	0.85	0.96		0.21	0.86	0.92	0.39	0.71	0.32
2,524,035	1950	0.68	0.96	0.85	0.75	0.45	0.89	0.66	0.93	0.90
2,543,181	1951	0.74	0.96		0.63	0.32	0.61	0.93	0.85	0.71
2,569,347	1951	0.72	0.96	0.79	0.63	0.48	0.72	0.79	0.99	0.95
2,642,679	1953	0.10	0.81		0.55	0.67	0.40	0.97	0.68	0.92
2,668,661	1954	1.00	0.87	0.80	0.98	0.85	0.83	0.01	0.01	0.01
2,682,050	1954	0.84	0.42		0.77	0.99	0.18	0.29	0.66	0.46
2,682,235	1954	0.88	0.63		0.60	0.99		0.10	0.23	0.15

Continued on next page

Ex ante measureable:		Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Patent	Grant Year	RETech	Cites	KPSS	KPST	Brdth	Orig	Priv Simm	LI Simm	Frgn Simm
2,691,028	1954	0.90	0.40		0.96	0.05		0.97	0.92	0.94
2,699,054	1955	0.21	0.97		0.97	0.23	0.77	0.98	0.97	0.98
2,708,656	1955	1.00	0.97		0.99	0.93		0.01	0.01	0.01
2,708,722	1955	0.91	0.97		0.78	0.44	0.99	0.08	0.75	0.43
2,717,437	1955	0.52	0.80		0.43	0.45	0.17	0.03	0.06	0.11
2,724,711	1955	0.99	0.39		0.82	0.06		0.46	0.75	0.78
2,752,339	1956	0.79	0.91		0.88	0.04	0.20	0.95	0.95	0.96
2,756,226	1956	0.89	0.60		0.71	0.18	0.98	0.34	0.63	0.86
2,797,183	1957	1.00	0.80		0.90	0.24	0.98	0.84	0.70	0.87
2,816,721	1957	0.59	0.95		0.72	0.46		0.42	0.52	0.72
2,817,025	1957	0.24	0.97		0.71	0.48		0.90	1.00	0.77
2,835,548	1958	0.75	0.79		0.85	1.00	0.97	0.65	0.42	0.55
2,866,012	1958	0.25	0.98		0.81	0.54	0.15	0.97	0.97	0.87
2,879,439	1959	0.72	0.97		0.77	0.80	0.55	0.44	0.82	0.55
2,929,922	1960	0.91	0.97	0.90	0.89	0.61		0.39	0.66	0.58
2,937,186	1960	0.99	0.58		0.89	0.12	0.13	0.99	0.97	1.00
2,947,611	1960	0.47	0.34	0.58	0.77	0.57		0.94	0.95	0.94
2,956,114	1960	0.45	0.90	0.71	0.74	0.55	0.39	0.98	0.98	0.90
2,981,877	1961	0.82	0.98		0.98	0.42	0.10	0.60	0.81	0.47
3,057,356	1962	0.89	0.97		0.93	0.54	0.09	0.55	0.94	0.58
3,093,346	1963	0.92	0.98		0.93	0.82	0.82	0.40	0.53	0.51
3,097,366	1963	0.28	0.55		0.41	0.99	0.73	0.86	0.62	0.77
3,118,022	1964	0.33	0.29	0.89	0.70	0.71		0.70	0.77	0.60
3,156,523	1964	0.05	0.46		0.85	0.25	1.00	0.86	0.83	0.98
3,174,267	1965	0.42	0.89	0.48	0.55	0.94	0.73	0.45	0.09	0.16
3,220,816	1965	0.06	0.54		0.85	0.42	0.09	0.03	0.14	0.09
3,287,323	1966	0.29	0.32	0.56	0.70	0.13	0.39	0.56	0.98	0.99
3,478,216	1969	0.42	0.80		0.84	0.37	0.62	0.35	0.64	0.46
3,574,791	1971	0.21	0.96	0.93	0.82	0.19		0.79	1.00	0.99
3,663,762	1972	0.38	0.97	0.84	0.78	0.68	0.26	0.13	0.40	0.09
3,789,832	1974	0.76	0.89		0.74	0.91	0.79	0.54	0.85	0.65
3,858,232	1974	0.32	0.98	0.97	0.71	0.78	0.95	0.58	0.94	0.77
3,906,166	1975	0.86	0.92	0.55	0.71	0.93	0.29	0.45	0.73	0.42
4,136,359	1979	0.71	0.76		0.97	0.69		0.63	0.98	0.83
4,229,761	1980	0.64	0.30		0.92	0.80	1.00	0.02	0.17	0.09
4,237,224	1980	0.96	0.98		1.00	0.62		0.05	0.22	0.14
4,371,752	1983	0.96	0.99		0.94	0.30	0.28	0.72	0.99	0.82
4,399,216	1983	1.00	0.99		1.00	0.68	0.26	0.04	0.18	0.07
4,464,652	1984	0.39	0.99	0.90	0.89	0.86	0.78	0.95	0.82	0.80
4,468,464	1984	1.00	0.92		1.00	0.58		0.04	0.16	0.12
4,590,598	1986	0.67	0.28	0.93	0.58	0.78	0.91	0.82	0.93	0.84
4,634,665	1987	0.96	0.78		0.99	0.53		0.05	0.18	0.07
4,683,195	1987	0.93	0.11	0.42	0.97	0.72		0.34	0.63	0.55
4,736,866	1988	1.00	0.92		1.00	0.47	0.24	0.02	0.07	0.02
4,744,360	1988	0.80	0.83		0.91	0.68	0.18	0.55	0.64	0.57
4,799,258	1989	0.80	1.00		0.95	0.32	0.95	0.20	0.55	0.14
4,816,397	1989	1.00	0.94		0.98	0.60	0.99	0.20	0.41	0.23
4,816,567	1989	0.94	0.15	0.82	0.99	0.66	0.58	0.27	0.54	0.34

Continued on next page

Ex ante measureable:		Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Patent	Grant Year	RETech	Cites	KPSS	KPST	Brdth	Orig	Priv Simm	LI Simm	Frgn Simm
4,838,644	1989	0.78	0.88		0.92	0.95	0.20	0.82	0.93	0.87
4,889,818	1989	0.92	0.99	0.53	0.98	0.71	0.20	0.50	0.75	0.65
4,965,188	1990	0.93	0.99	0.49	0.98	0.75	0.20	0.40	0.72	0.61
5,061,620	1991	0.93	0.99		1.00	0.36		0.11	0.16	0.10
5,071,161	1991	0.48	1.00		0.67	0.94	0.20	0.68	0.47	0.39
5,108,388	1992	0.91	0.88	0.42	0.97	0.85	0.45	0.42	0.54	0.36
5,149,636	1992	0.66	0.48		0.99	0.52	0.19	0.05	0.18	0.07
5,179,017	1993	0.95	0.41		1.00	0.31		0.16	0.33	0.24
5,184,830	1993	0.62	0.94		0.98	0.34	0.43	0.84	0.88	0.89
5,194,299	1993	0.51	0.94	0.81	0.73	0.31	0.81	0.62	0.48	0.56
5,225,539	1993	0.98	0.99		1.00	0.52		0.10	0.26	0.17
5,272,628	1993	1.00	0.99	0.99	0.99	0.23	0.78	0.14	0.44	0.13
5,747,282	1998	1.00	0.45	0.08	0.97	0.15		0.39	0.30	0.32
5,770,429	1998	0.99	0.01		0.61	0.13	0.86	0.36	0.27	0.31
5,837,492	1998	1.00	0.01	0.04	0.83	0.28		0.39	0.39	0.40
5,939,598	1999	1.00	0.57	0.53	1.00	0.31	0.68	0.21	0.28	0.27
5,960,411	1999	0.87	1.00	1.00	1.00	0.03	0.85	0.18	0.50	0.09
6,230,409	2001	0.53	0.07		0.75	0.96	0.74	0.80	0.24	0.38
6,285,999	2001	0.92	1.00		0.99	0.13	0.74	0.13	0.47	0.12
6,331,415	2001	0.65	0.97	0.95	0.99	0.60	0.47	0.39	0.64	0.49
6,455,275	2002	0.99	0.45		0.98	0.13	0.12	0.45	0.40	0.46
6,574,628	2003	0.87	0.98		1.00	0.04	0.74	0.29	0.65	0.22
6,955,484	2005	0.07	0.75		0.78	0.68	0.63	0.63	0.24	0.35
6,985,922	2006	0.83	0.98		0.93	0.10	0.90	0.21	0.62	0.11

Table A4: Determinants of Startups' Financing

This table presents cross-sectional tests relating startups' ex ante technological traits and their financing through new VC rounds and syndicated debt issues. The sample and controls are the same as in Table VII. For brevity, we only report the main coefficient on *RETech*. Columns (1) and (3) equal 100 when the firm obtains debt in quarter q and zero otherwise. Columns (2) and (4) equal are log of one plus the dollars raised through funding in quarter q. Data on debt issuance comes from DealScan. To facilitate interpretation, *RETech* is standardized. All independent variables are lagged one quarter. Adjusted R^2 is reported as a percentage. Standard errors are clustered by startup and t-statistics are reported in parentheses. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	New VC Round		New Bank Debt	
	Dummy (x100)	Log Size	Dummy (x100)	Log Size
	(1)	(2)	(3)	(4)
RETech	0.347***	0.026***	0.032*	0.005*
	(3.81)	(3.88)	(1.94)	(1.83)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm Age FE	Yes	Yes	Yes	Yes
Firm Cohort FE	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	347,918	347,918	347,918	347,918
R2 (%)	3.3	3.8	1.5	1.6

Internet Appendix for

"Rapidly Evolving Technologies and Startup Exits"

August 2021

This appendix contains additional material not reported in the paper to preserve space.

IA.A Defining the Entity Type of Patents' Assignees

To classify if a patent is granted to (A) a private, domestic U.S. firm, (B) an international firm, or (C) a U.S. public firm, we use the following procedure. First, we find all patents assigned to public firms. We obtain the GVKEY for assignees from the NBER patent dataset and augment this with data from Kogan, Papanikolaou, Seru, and Stoffman (2016). We use all assignee links for the entire 1900-2013 period. Also, note that Kogan, Papanikolaou, Seru, and Stoffman (2016) contains PERMNO identifiers, which we convert to GVKEY using a link table from WRDS. When the headquarters country from CRSP-Compustat is available, we mark these firms as either international firms or U.S. public firms. Next, we output the top 3,000 remaining assignees and manually classify the entity type. After these steps, 3,126,605 patents are classified as either U.S. public firms or foreign firms.

Second, we use information from the NBER classification of assignees and manual categorization to remove patents assigned to governmental entities, research think tanks, or universities.

Third, we directly identify patents assigned to foreign firms when the last word in the assignee name is an unambiguous foreign legal identifier, such as "GMBH", "PLC", and "Aktiengesellschaft". We also identify patents granted to foreign firms when the assignee is a firm (e.g. "CORP") and USPTO data indicates that the assignee is not domestic. This step identifies 898,797 patents granted to foreign firms.

Fourth, we classify entities as U.S. private domestic firms when the assignee is a firm (e.g. "CORP") and USPTO data indicate the assignee is domestic. Previous steps affirmatively prevent us from calling a corporation a private domestic firm if the corporate is a public firm, a think tank, or an international corporation.

In total, we classify the entity type of 78% of all patents granted from 1900-2013. Moreover, during our main analysis period (1980-2010), we can classify the assignee entity type for 92% of patent applications. Of the 4,161,306 applied for in the main analysis period, 12% are private U.S. firms, 27% are public U.S. firms, 41% are foreign firms, 8% are unclassified, and 11% are "other".

For Figure IV, we additionally identify a subset of the 11% of patents in the "other" classification as individuals. The primary component of this step is to define an assignee as an individual if (a) the assignee string has no common firm terms (e.g. "AAA" or "INC"), (b) the string has a common name structure (e.g. First, Middle Initial, Last), and (c) the string uses a first name above the median in the census name file.[47] This rule has a false positive rate of just 2% in tests and a false negative rate of 36%. To catch some additional false negatives, we look for common first and last name combinations, and augment this with a final manual pass.[48]

[47] We combine the rankings for the 1970 and 1990 birth cohorts from `https://www.ssa.gov/OACT/babynames/limits.html`.

[48] We use the top 2,000 surnames from `https://www.census.gov/topics/population/genealogy/data/2000_surnames.html` and the top 1,500 first names from the 1970 and 1990 birth cohorts.

IA.B Matching patents to VentureXpert

We download all data on firms receiving venture capital funding starting in 1970 and ending in 2013 from VentureXpert using SDC Platinum. In addition to the dates of venture financing, we also download data indicating each portfolio company's founding date, resolution type (as IPO, acquisition, or unresolved) and date, the company's name, and the number of financing rounds it received.

Merging VentureXpert with the patent-level data requires a link between firms in the patent database (the initial assignees) and firms in the VentureXpert database. We develop a fuzzy matching algorithm—outlined below—to match firms in both databases using their names. The algorithm matches 532,660 patents granted between 1966 and 2013 to 19,324 VC-backed firms.[49] 96.6% of the patent matches and 90.7% of the VC-backed firms are matched via exact matches on the raw firm name in both datasets or on a cleaned version of the firm name.

The matching procedure begins by standardizing assignee names in the patent dataset and in Venture VentureXpert, using a name standardization routine from Nada Wasi.[50] This standardizes common company suffixes and prefixes and produces stem names. We also modify this program to exclude all information after a company suffix, as this is typically address information erroneously stored in the name field by the USPTO. After standardizing the names, we use the following steps to match firms in the two datasets:

1. We compare all *original string* names in each dataset, adjusted only to replace all uppercase characters. If a single VC-backed firm is an exact match where the patent application is after the firm's founding date, we accept the match. This step matches 59,026 patents to VC-backed firms (11% of the accepted matches).
2. For the remaining patents, we compare all *cleaned string* names in each dataset. If a single VC-backed firm is an exact match where the patent application is after the firm's founding date, we accept the match. This step matches 455,456 patents to VC-backed firms (86% of accepted matches).
3. For the remaining patents, we select matches using a fuzzy matching technique, with rules based on random sampling and validation checks in a hold out sample. This step matches 18,178 patents to VC-backed firms (3% of accepted matches). The steps are as follows:
    (a) We compute string comparison scores by comparing all *cleaned string* names in each dataset using several different string comparison functions. We do this three separate times, requiring that (1) the first three characters are exact matches, (2) the first five characters are exact matches, and (3) the first seven characters are exact matches. We then output a random sample of patents for an RA to examine.
    (b) The highest performing rule was a bi-gram match function with the restriction that the first seven characters were equivalent in both the patent assignee and company name. For each remaining patent, we keep as *possible* matches any pair with equal name stems (the first seven characters) and the highest bi-gram match above 75%.
    (c) A random subset of singleton possible matches, in addition to all borderline suggested matches, were reviewed by hand.

As a result of this matching process, our patent-level database contains U.S. private firms that both (A) have patents and (B) have received VC funding. Aside from imperfections in the matching process, which could be material, this database is the universe of such firms.[51] For each such firm, we have data indicating its final outcome and text-based data indicating the details of the firm's patents, and when

[49] Firms can receive patents before VC funding.

[50] http://www-personal.umich.edu/∼nwasi/programs.html

[51] Lerner and Seru (2017) note that using string matching to identify firms suffers from a limitation when private firms have patents issued to legal entities with different names, such as subsidiaries or shell

they were applied for and granted. This data allows us to examine both (A) potential drivers of VC funding among firms that have patents but have not yet received funding, and (B) final resolutions of private status as IPOs or acquisitions. Cross-sectional and time-series examination of both form the basis of our hypothesis testing.

companies meant to obfuscate the owner. This limitation can not be avoided but is reduced for our sample of interest. VC-backed private firms are typically small and thus are unlikely to have distinctly named subsidiaries. Moreover, obfuscation is most often used by "non-practicing entities", often called patent trolls, which are unlikely to be a material number of firms in our sample of VC-backed startups.

Figure IA.1: Trends in Aggregate Technology Variables
This figure reports characteristics of the aggregate patent corpus from 1930 to 2010. The variables are defined at the patent level in Section II. The aggregate stocks reported below are the average of each variable across all patents for the prior 20 quarters after applying a 5% quarterly rate of depreciation. The series presented are four quarter moving averages to smooth out seasonality.

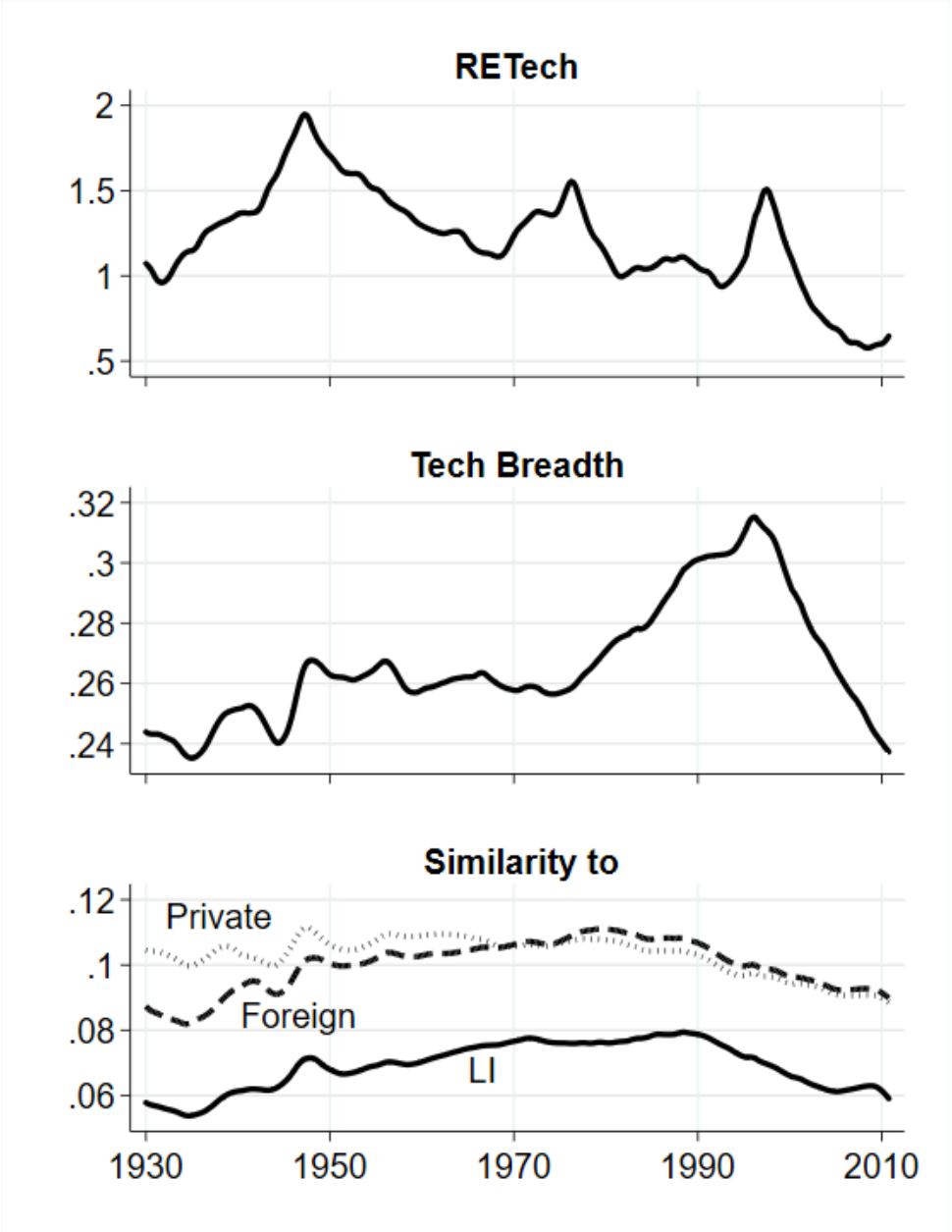


Figure IA.2: Counterfactual: Holding the Patent Wordspace Fixed

This figure reports the evolution of *RETech* for the aggregate patent corpus from 1930 to 2010 as computed in Figure II. We then compute a counterfactual RETech for all patents applied for in 1960 or later as

$$\text{Counterfactual RETech}_{j,t} = \left(\frac{\tilde{B}_{j,t}}{\tilde{B}_{j,t} \cdot \mathbf{1}} \cdot \tilde{\Delta}_t \right) \times 100$$

where $\tilde{\Delta}_t$ is 492,240 randomly drawn words from Δ_t. This ensures that the size of the wordspace from 1960 on remains constant ($\tilde{B}_{j,t}$ is the corresponding subset of $B_{j,t}$). The aggregate stocks reported below are the average RETech and *Counterfactual RETech* across all patents for the prior 20 quarters after applying a 5% quarterly rate of depreciation. All series are reported as four quarter moving averages.

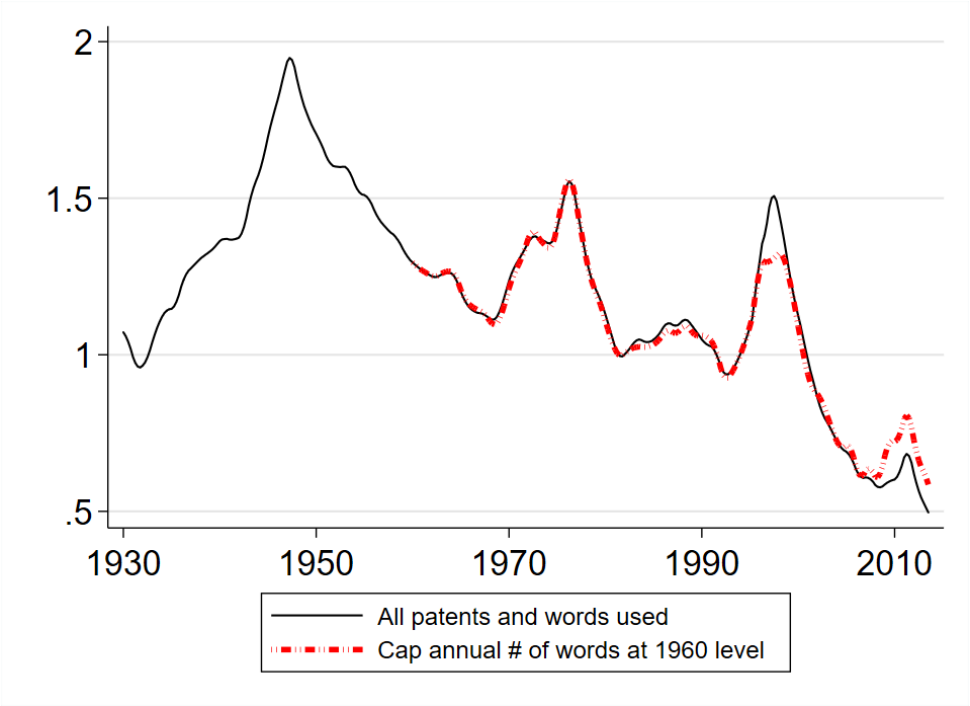


Figure IA.3: RETech After Stripping Fixed Effects

This figure repeats the plot of *RETech* for the aggregate patent corpus from 1930 to 2010 from Figure II in the solid black line. That figure converts patent-level *RETech* to a time series as the average RETech across all patents for the prior 20 quarters after applying a 5% quarterly rate of depreciation. Here, we run patent-level regressions $RETech_{j,c,i} = \eta_c + \epsilon_{j,i,c}$ and $RETech_{j,c,i,s,L} = \eta_c + \eta_i + \eta_s + \eta_L + \epsilon_{j,i,c}$ where patent j is in 2-digit NBER technology category c, assigned to assignee i from state s and was filed by lawyer L. We convert the residuals of these regressions to aggregate stocks using the same function as *RETech* and plot them below. The technology-only regression is plotted as hash marks, and the technology plus assignee regression is plotted as dashes. Data on state plus identifiers for lawyers and assignees are obtained from PatentsView and are available for patents granted after 1976. We begin the plot for that regression in 1980 after a burn-in period of 4 years. All series are reported as four quarter moving averages.

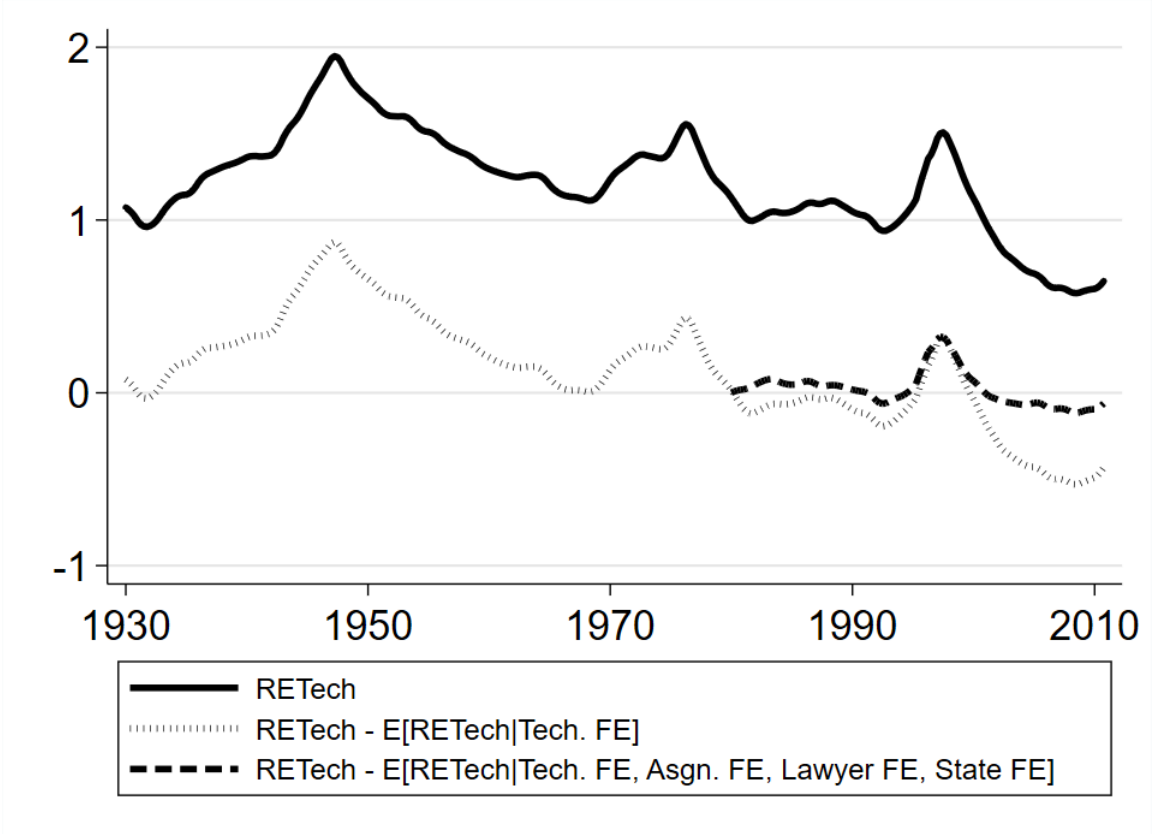
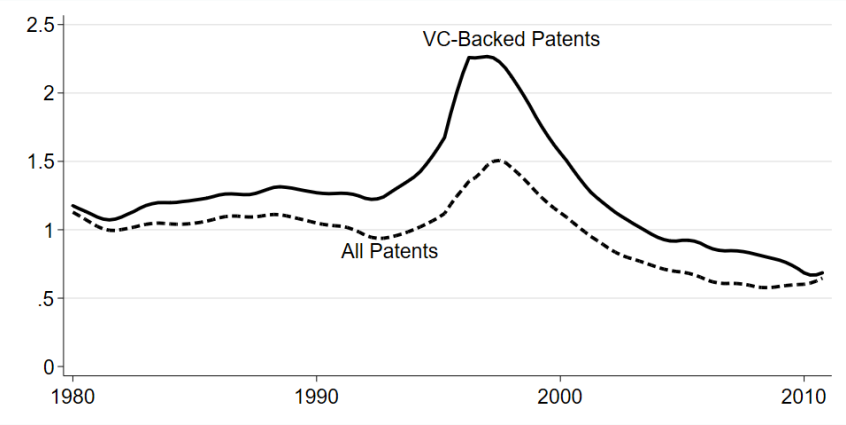


Figure IA.4: VC-Backed Startups: RETech and Exit Trends

This figure compares trends among the startup sample to aggregate data. Panel A reports *RETech* based on patents held by VC-backed startups from 1980 to 2010 (solid line) and across all patents (dashed line). *RETech* is defined at the patent level in Section II. The time series are constructed from the patent-level data as stocks, following the same procedure as in Figure II. Panel B reports in the solid lines the percentage of startups that exit in the sample during the year via IPO or sell-out (left axis). The dashed lines report aggregate trends on IPOs and sell-outs of private targets and are reported in dashed lines as a fraction of lagged real GDP (right axis). Real GDP is in units of $100m. We obtain data on aggregate IPOs from Jay Ritter's website, and exclude non-operating companies, as well as IPOs with an offer price lower than $5 per share, unit offers, small best effort offers, bank and savings and loans IPOs, natural resource limited partnerships, companies not listed in CRSP within 6 month of their IPO, and foreign firms' IPOs. Data on acquisitions are from the Thomson Reuters SDC Platinum Database, and include all domestic, completed acquisitions of private firms coded as a merger, acquisition of majority interest, or acquisition of assets giving the acquirer a majority stake. All series are reported as four quarter moving averages.

Panel A: RETech – Startup Sample vs. Aggregate Data



Panel B: Exits – Startup Sample vs. Aggregate Data

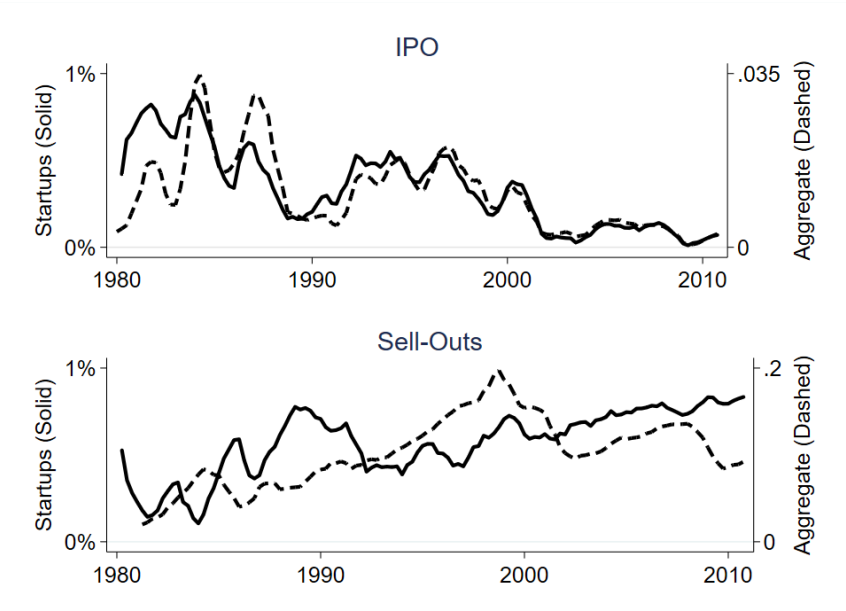


Table IA.1: Robustness of Baseline Results to Aggregration of RETech

This table presents robustness tests of the main results in Table VII. For brevity, we only report the main coefficient on *RETech* for each test. Each row (Test #) corresponds to an alteration of the main test, wherein we recompute RETech for a given startup-quarter using a different aggregation function. Except for the listed alteration, each of the models within a row repeats the corresponding model in the same column of Table VII. To facilitate interpretation, competing risk models report exponentiated coefficients minus one, OLS models report coefficients scaled by 100, and *RETech* is standardized and lagged one quarter. Standard errors are clustered by startup unless otherwise noted and t-stats are reported in parentheses. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

		Competing Risk Hazard		OLS (coefficients x100	
		IPO	Sell-Out	IPO	Sell-Out
Test #	RETech definition	(1)	(2)	(3)	(4)
(1)	Average across patents in last year	0.183***	-0.209***	0.071***	-0.147***
		(10.49)	(-9.34)	(3.95)	(-9.55)
(2)	Average across patents in last 2 years	0.228***	-0.184***	0.079***	-0.130***
		(12.85)	(-7.72)	(4.47)	(-6.89)
(3)	Average across patents in last 3 years	0.223***	-0.110***	0.063***	-0.067***
		(11.52)	(-4.82)	(3.68)	(-3.20)
(4)	Average across patents in last 4 years	0.216***	-0.063***	0.049***	-0.010
		(10.10)	(-2.91)	(2.79)	(-0.45)
(5)	Average across patents in last 5 years	0.209***	-0.046**	0.041**	0.029
		(9.04)	(-2.13)	(2.27)	(1.29)
(6)	Max across patents in last year	0.195***	-0.251***	0.098***	-0.164***
		(12.06)	(-9.49)	(4.78)	(-9.76)
(7)	Max across patents in last 2 years	0.234***	-0.228***	0.101***	-0.161***
		(12.68)	(-8.50)	(4.68)	(-7.92)
(8)	Max across patents in last 3 years	0.224***	-0.139***	0.082***	-0.090***
		(10.76)	(-5.56)	(3.80)	(-3.97)
(9)	Max across patents in last 4 years	0.215***	-0.073***	0.059***	-0.012
		(9.41)	(-3.19)	(2.70)	(-0.49)
(10)	Max across patents in last 5 years	0.207***	-0.052**	0.050**	0.032
		(8.49)	(-2.29)	(2.19)	(1.27)

Table IA.2: The Determinants of Startups' Exits - Lawyer Fixed Effects

This table presents cross-sectional tests relating startups' ex ante technological traits to their exit. Each of the models repeats the corresponding model from Table VII, but replaces the main variable *RETech* with a version that has been orthogonalized with respect to lawyer fixed effects. That is, we first run patent-level regressions $RETech_{j,i} = \eta_i + \epsilon_{j,i}$ where lawyer i is associated with patent j. We add the mean value of RETech back to the fitted residuals (i.e., at the patent level, $\widetilde{RETech} \equiv \widehat{\epsilon_{j,i}} + \overline{RETech_{j,i}}$) and finally aggregate $\widetilde{RETech}$ to the startup-quarter with the same procedure as for *RETech*. For brevity, we only report the coefficients on $\widetilde{RETech}$. To facilitate interpretation, competing risk models report exponentiated coefficients minus one, OLS models report coefficients scaled by 100, and $\widetilde{RETech}$ is standardized and lagged one quarter. Independent variables are lagged one quarter. Adjusted R^2 is reported as a percentage. Standard errors are clustered by startup and t-stats are reported in parentheses. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Competing Risk Hazard		OLS	
	IPO	Sell-Out	IPO	Sell-Out
	(1)	(2)	(3)	(4)
$\widetilde{RETech}$	0.199***	-0.109***	0.049***	-0.055***
	(9.50)	(-4.71)	(2.82)	(-3.08)
Controls	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes
Industry FE	No	No	Yes	Yes
Technology FE	No	No	Yes	Yes
Location FE	No	No	Yes	Yes
Firm Age FE	No	No	Yes	Yes
Firm Cohort FE	No	No	Yes	Yes
Observations	346,490	345,403	342,146	342,146
R2 (%)	N/A	N/A	0.7	0.9

Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch