Technological Disruptiveness and the Evolution of IPOs and Sell-Outs

Donald Bowen, Laurent Frésard, and Gerard Hoberg*

September 19, 2018

ABSTRACT

We show that the recent decline in IPOs in U.S. markets is explained by changes in the technological disruptiveness of startups, which we measure using textual analysis of patents from 1930 to 2010. We focus on startups backed by venture capital and show that startups with disruptive technologies are more likely to exit via IPO and are less likely to exit via sell-out. This is consistent with IPOs being favored by firms with the potential to carve out independent market positions with strong defenses against rivals. We document an economy-wide trend of declining technological disruptiveness since World War II that accelerated since the late 1990s. These trends predict fewer IPOs and more sell-outs, and we find that 20% to 60% of the recent dearth of IPOs, and 55% of the surge in sell-outs, can be attributed to changes in firms' technological characteristics.

Key words: Initial Public Offerings (IPOs), Acquisitions, Sell-Outs, Technology, Disruptiveness, Venture Capital

JEL classification: G32, G34, G24

---

*Virginia Tech, the University of Lugano, Swiss Finance Institute, and University of Maryland, and the University of Southern California respectively. Bowen can be reached at donbowen@vt.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. For helpful comments, we thank Adrien Matray, Francesco D'Acunto, Francois Degeorge, Francois Derrien, Francesco Franzoni, Xuan Tian and seminar participants at Bocconi University, the CCF conference in Bergen, the SFI Research days, Universita Cattolica, Texas Christian University, Tsinghua University, CKGSB, the University of Lugano, the University of Zurich, and Virginia Tech. All errors are the authors' alone.

I Introduction

Since the late 1990s, the number of initial public offerings (IPOs) in U.S. markets has sharply declined. Over the same period, the number of private firms exiting via acquisitions (i.e., trade sales) has soared. Successful startups are nowadays more likely to sell out to other (public or private) companies than seek independent public listings. Many observers in the media and policy circles worry that these trends reflect a general erosion in the ability of U.S. financial markets to spur economic growth and spread its benefits across the general public.[1] Recent studies indicate, however, that the dearth of IPOs is unlikely due to regulatory changes affecting public firms (Gao, Ritter, and Zhu (2013) and Doidge, Karolyi, and Stulz (2013)) and is partly explained by changes in regulations affecting the financing ability of private firms (Ewens and Farre-Mensa (2018)). Understanding the underlying mechanisms driving the decline in IPOs remains of paramount importance as the assessment of competing policy responses depends critically on this microfoundation.

In contrast to existing regulation-based explanations, we show that the recent shift from IPOs to trade sales is strongly related to a decline in firms' technological disruptiveness. Our analysis builds on the long-standing idea that exiting through an IPO or by selling out to another firm depends on which exit type enhances the growth potential of successful private firms (Bayar and Chemmanur (2011) or Gao, Ritter, and Zhu (2013)). Accordingly, we predict that private firms that can develop an independent market presence by creating new markets or disrupting existing ones with patented technologies are more likely to exit and scale up through a public listing.[2] In contrast, firms with less disruptive technology, such as technology that can improve existing products, are more likely to sell out and thus expand within the boundaries of existing firms that can more efficiently scale these technologies. We show that the overall technological disruptiveness

[1] See for instance "The endangered public company: The big engine that couldn't," *The Economist* (May 19, 2012) or "US stock markets seek depth in IPO pool," *Financial Times* (January 9, 2018).

[2] Consistent with this idea, Darby and Zucker (2018) show that biotechnology firms go public when they have a science base that can be successfully commercialized, Chemmanur, He, He, and Nandy (2018) report that manufacturing firms are more likely to go public than sell out when they already have a strong product market presence (i.e., market share), Poulsen and Stegemoller (2008) and Cumming and Macintosh (2003) show that firms with more growth potential favor exit through IPOs.

of U.S. firms has significantly decreased in recent years, leading to fewer IPOs and more trade sales.

We study the exit strategies of a large sample of U.S. startups backed by venture capitalists (VCs), because detailed data enables us to precisely link startups' choice of exits (i.e., IPO or sell-out) to the patents they file with the U.S. Patent and Trademark Office (USPTO).[3] We use the text in patents to develop a novel measure of technological disruptiveness, that is, a patent's ability to create new markets or generate radical changes in existing ones (Abernathy and Utterback (1978)). Specifically, we define the technological disruptiveness of a given patent by the intensity with which its text contains vocabulary that is new and fast-growing compared to existing knowledge. For example, genetics-related words such as "peptide", "clone", or "recombinant" are newly and increasingly appearing in patents in 1995. We thus define patents that extensively employ such words in 1995 as more disruptive.[4] As they contain pioneering and fast growing ideas that are legally protected, firms holding more disruptive patents are likely to radically change markets and capture valuable competitive advantages.[5] We posit that these firms are more likely to go public.

Consistent with recent evidence suggesting that new innovative ideas are getting harder to discover and develop (Jones (2009) and Bloom, Jones, Reenen, and Webb (2017)), the average disruptiveness of U.S. patents has markedly declined since 1950, except for temporary spikes during the 1970s (i.e., computer revolution) and the 1990s (i.e., the internet revolution). A simple comparison of the aggregate evolution of our measure of technological disruptiveness against that of private firms' exits over the last four decades strongly suggests that IPOs are more prevalent when innovation is more disruptive, whereas trade sales prevail when innovation is less disruptive.

We confirm this interpretation at the micro-level by exploring whether the technolog-

[3] In addition, VC-backed startups account for a large fraction of the IPO and acquisition market (Ritter (2017)) and produce a substantial share of innovation in the economy (Gornall and Strebulaev (2015)).

[4] With our text-based approach, we classify unambiguously important inventions such as the jet engine, transistor, laser, satellite, or more recently PageRank (Google) as highly disruptive.

[5] In line with this conjecture, we show that disruptive patents attract significantly more citations and higher stock market valuation upon publication by the USPTO.

ical disruptiveness of VC-backed startups explains the choice of exit. We focus on VC-backed startups that receive at least one patent over the 1980-2010 period and compute the technological disruptiveness for each startup-quarter based on all patents received in the previous five years. We thus obtain a large panel of 561,982 startup-quarter observations representing 13,679 distinct startups and 506,096 patents. Using a competing risks regression model that accounts for possible interdependences between exits (i.e., IPO or sale), we find strong evidence that startups with more disruptive patents are significantly more likely to go public and less likely to sell out. The link between technological disruptiveness and startup exits is economically large, as a one standard deviation increase in a startup's technological disruptiveness is associated with a 21.8% increase in its IPO rate and a 17.9% decrease in its rate of exit through trade sales.

These results remain after controlling for startups' age, size, existing characteristics of patent portfolios (e.g. citations and originality), financing rounds, and overall financial market conditions. They also hold in specifications with fixed effects that absorb variation within startup cohorts, geographic locations, and technological expertise.[6] Furthermore, we provide additional evidence supporting the validity of our new measure of startup technological disruptiveness and its role in explaining IPO rates. For example, we delineate each startup's product market based on the textual similarity between its business description and the 10-K business descriptions of publicly traded firms. As direct validation of our measure, we find that the publicly traded firms that are most similar to a given startup discuss market disruption significantly more when that startup displays high level of technological disruptiveness. Alternatively, we show that among startups exiting through IPOs, those with more disruptive technologies exit into less stable and more competitive markets that intuitively offer greater opportunities for disruption.

We also find that startups' exit choices are related to other novel text-based technological measures. For instance, we define the technological "breadth" of each patent as the intensity with which its text combines vocabulary from diverse bodies of knowledge

[6]Our results are also robust to changes in econometric specifications that vary the horizon over which we measure exits (ranging from the next quarter to the next five years), and to focusing on the early part of the sample to limit potential truncation bias relating to startup resolution and patent grants (Lerner and Seru (2017)).

(as opposed to specialized knowledge). Startups with more technological breadth are more likely to exit via IPO and less likely to sell out, consistent with the idea that less specialized technology is less redeployable toward other uses. We also construct various measures of technological "similarity" by directly comparing a patent's vocabulary to the vocabulary used by different groups of firms. In line with Phillips and Zhdanov (2013) and Wang (2017), we show that startups with patents that are more similar to that of lead innovators (e.g., IBM, Microsoft, or Intel in the 2000s) are more likely to sell out. In contrast, sell-outs are less likely for startups having patents that are more similar to that of other private or foreign firms, consistent with the notion that a firm is less likely to be an acquisition target when there are more alternative targets.

Given our finding that technological traits such as disruptiveness explain the choice of exit as IPO or trade sale in the cross-section, it is natural to examine the economy-wide time-series properties of these variables. The central question is whether dramatic changes in technological traits can help to explain the dearth of IPOs and the surge in trade sales observed in recent years. We thus proceed with a two step procedure inspired by the disappearing dividends literature.[7] We first estimate our cross-sectional exit models over an initial period 1980-1995 and predict average IPO and sell-out incidence rates over the 1996-2010 period. Second, we compute average predicted values from this model in an out-of-sample period (1996-2010) and compare the average model's predicted rate of IPO exits to the average actual rate of IPO exits in this later period.

When our technological characteristic variables are excluded from the model, the average predicted IPO incidence is 0.63% per quarter, whereas the actual IPO incidence rate is 0.34%, confirming the existence of the disappearing IPOs anomaly. When we including our technological characteristics, the predicted IPO incidence rate decreases materially to 0.57%. Furthermore, the explanatory power of these technological changes is not sensitive to the periods used to estimate and predict the models, or the lags between the measurement of startup technological traits and predicted exits. We conclude, across sev-

[7] This methodology was used by Fama and French (2001) and Hoberg and Prabhala (2009) to explain disappearing dividends, and more recently by Bates, Kahle, and Stulz (2009) to explain increasing cash holdings.

eral specifications, that changes in technological characteristics explain roughly 20% of the overall decline in IPO rates. A similar analysis reveals that changes in technological traits explain roughly 55% of the recent rise of trade sales.

We further propose that the ability to create disruptive technologies is not uniform across industries, as the difficulty of creating disruptive inventions is likely convex as industries mature.[8] A consequence is that any residual ability to create new inventions should vanish at increasing rates as industries mature, and initial industry conditions regarding the product life cycle become crucial in predicting the future IPO rate. We consider the product life cycle in Abernathy and Utterback (1978) and posit that markets in transition or those reaching maturity (e.g., markets that have reached a dominant product design) are likely to experience the most extreme decline in IPO rates. On the other hand, very young "fluid" markets are more likely to have high IPO rates that likely grow during the earliest years as competing disruptive product designs and inventions require some initial "time to build".

We follow Hoberg, Phillips, and Prabhala (2014) and compute the degree of product market fluidity in each firm's product market. For our VC-backed startups, we use the business description text, which is observable at the time of "first money". We find strong evidence that IPO rates do decline faster as product markets mature as our technological characteristics explain an economically large 60% of the decline in IPOs in markets with below median product stability. In contrast, we only explain 10% of the decline in fluid product markets. Looking to the future, the IPO rate will likely depend critically on the extent to which new fluid industries emerge (such as those based on artificial intelligence, fintech, and other new technologies) and the extent to which existing industries stabilize. Established industries, in particular, should experience persistently declining IPO rates.

Our analysis adds to a growing literature that considers changing market conditions and the link to the disappearance of IPOs and the rise of trade sales.[9] Ewens and Farre-

[8] The intuition for this assumption is that the most logical inventions are discovered first. As industries mature, the best ideas become "picked over" and later inventions arise only with very high search costs.

[9] Ritter and Welch (2002) and Lowry, Michaely, and Volkova (2017) provide comprehensive surveys of the literature on IPOs.

Mensa (2018) suggest that the IPO decline results from the increased bargaining power of founders (over investors), their preference for control, and inexpensive capital in the private market. Gao, Ritter, and Zhu (2013) suggest that the decline in IPO originates from changes in market structure that favor selling out to realize economies of scope. Doidge, Kahle, Karolyi, and Stulz (2018) argue that an increased focus on intangibles also plays an important role. Our paper shows that changes in technological traits (especially disruptiveness), which strongly explain exit methods in cross section, can also explain the decline in IPOs and the rise of trade sales in time series. To our knowledge, our paper is the first to quantify how much of the observed trends are attributable to changes in firms' technological characteristics. Although more work is needed to draw clear policy recommendations, our findings suggest that the decline in IPOs might be particularly unresponsive to policy changes, with subsidies to innovative entrepreneurs being perhaps a possible exception.

Our findings also add to the literature studying the determinants of startup exits. Existing research indicates that the exit choice of startups depends on founders' private benefits of control, product market presence, and growth potential (Cumming and Macintosh (2003), Bayar and Chemmanur (2011), Poulsen and Stegemoller (2008) or, Chemmanur, He, He, and Nandy (2018)). Our analysis emphasizes the importance of technological characteristics including disruptiveness, breadth, and similarity to economically relevant firms.

II Text-based Technological Characteristics

In this section, we describe the patent data and explain the construction of our new measures of technological characteristics based on patent text.

A Patent Text

We use a web-crawling algorithm to gather information from Google Patents for all 6,850,075 patents that were applied for between 1930 and 2010 and granted by 2013. For each patent, we gather the publication date, application date, names 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). Since we are interested in measuring the technological changes pertaining to the corporate sector, we categorize each patent into groups based on four types of applicants: U.S. public firms, U.S private firms, foreign (private or public) firms, or others (e.g., universities or foundations). For brevity, we describe this classification method in Appendix A.

[Insert Figure I about here]

The full text of each patent consists of three distinct sections: abstract, claims, and patent description. The claims section defines the scope of legal protection granted. The description section explicitly describes the characteristics of the invention/innovation. It typically includes a title, technical field, background art, specification example, and industrial applicability. The abstract contains a summary of the disclosure contained in the description and claims sections. Figure I presents an example of a typical patent textual structure (#6285999, "A method for node ranking in a linked database", assigned to Google in 1998.). We append all three sections into a unified body of text because earlier patents do not include all sections, and because the organization of patent text into the three sections may have changed over time (Kelly, Papanikolaou, Seru, and Taddy (2018) and Packalen and Bhattacharya (2018)).[10]

Following earlier studies constructing variables from text (e.g., Hanley and Hoberg (2010) or Hoberg and Phillips (2016)), we represent the text of each patent as a numerical vector with a length equal to the number of distinct words in the union of all patent applications in a given year t. We denote this length N_t.[11] Following the convention in the literature, we eliminate commonly-used words (words appearing in more than 25%

[10] These papers also use patent text to identify major ideas from each historical year based on each new word's intensity of use in future patents, and use these major ideas to determine whether patents use old or new technologies. Our text-based measures differ in two important ways. First, we propose measures capturing distinct (and complementary) dimensions of patents' technological characteristics. Second, our measures can be measured ex-ante, that is, using only past and current-year patents. This distinction is important as our goal is to predict *future* outcomes (i.e., startups' exits) using initial technological characteristics without look-ahead bias.

[11] We organize patents based on their application year rather than the year of the patent grant, as this more accurately reflects the timing of innovation.

of all patents in a given year) and rare words (words appearing only in one patent in a given year).[12] Each patent j applied for in year t is then represented by a vector $V_{j,t}$ (of length N_t) in which each element corresponds to the number of times patent j employs one of the unique N_t words used in year t. If patent j does not use a given word, the corresponding element of $V_{j,t}$ is set to zero. This vectorization procedure insures that all patent applications in a given year are represented by a collection of vectors that are in the same space (of dimension N_t).

Due to the large number of words used across all patents in a given year, the vectors $V_{j,t}$ are quite sparse with most elements being zero. For instance, in 1980, the number of distinct words used in an average patent is 352, and the median is 300, while there are 400,097 distinct words used across all patent applications. In 2000, the average and median are 453 and 338, and the total across all applications is 1,358,694.

B Technological Disruptiveness

To capture the technological disruptiveness of a given patent, we focus on the extent to which a patent uses vocabulary that is new or experiencing rapid growth compared to existing knowledge. Intuitively, disruptive patents are based on ideas that radically change an industry or business strategy (e.g., new products or processes), either by creating a whole new market or by disrupting an existing one. We thus posit that disruptive patents use novel words that are fast-growing in the universe of patents.

To quantify disruptiveness, we rely on the rate of change in the use of each word among all patent applications during the current and prior year. As our goal is to link technological disruptiveness to future outcomes (i.e., startups' exits), we construct technological disruptiveness using only text contained in all past and contemporaneous patents. This alleviates forward-looking bias in our measure. We define an aggregate vector Z_t (of length N_t) in each year as having elements containing the number of times a given word is used across all patent applications in year t. This vector thus represents the aggregate

[12] Given the highly technical and rapidly evolving nature of text in the patent corpus, we do not implement additional filters (e.g. nouns only). While this choice might potentially introduce noise into our measurements, it maintains power.

frequency of word usage in a given year t. We then define the vector D_t as the annual rate of change in the usage of each word (from $t-1$ to t) as:[13]

$$D_t = \frac{Z_t - \tilde{Z}_{t-1}}{Z_t + \tilde{Z}_{t-1}}, \tag{1}$$

where division is element-by-element.

The set of annual vectors D_t thus tracks the appearance, disappearance, and growth of specific technological vocabulary over time. Elements of D_t are positive if the usage of the corresponding words increases from year $t-1$ to t, and negative if it decreases (e.g., words becoming obsolete).

[Insert Table I about here]

As an illustration, Table I displays the ten words experiencing the largest increases and decreases in use across all patent applications in specific years. For instance, in 1995, we detect an acceleration of terms related to genetics, such as "polypeptides", "clones", "recombinant" and "nucleic", following rapid progress in genome sequencing. In that year, use of terms such as "cassette," "ultrasonic," and "tape" are sharply decreasing. In 2005, the most rapidly growing words are related to the internet and include terms such as "broadband", "click", "configurable", or "telecommunications".

To obtain the level of technological disruptiveness for a given patent j, we take the frequency-weighted average of the vector D_t based on the words that patent j uses as follows:

$$\text{Tech Disruptiveness}_{j,t} = \frac{V_{j,t}}{V_{j,t} \cdot \mathbf{1}} \cdot D_t \times 100, \tag{2}$$

where the operator "$\cdot$" denotes the scalar product between two vectors, and "1" is a unit vector of dimension N_t. Intuitively, patents using words whose usage surges across all patent applications (i.e., have positive entries in the vector D_t) have higher levels of disruptiveness. This is the case for patents that either employ words that appear in the

[13] To ensure Z_t and Z_{t-1} are in the same space, we modify Z_{t-1} by adding zero elements for words that newly appeared in year t (as they were not originally in the $t-1$ space). Analogously, we modify Z_t by adding zero elements for words that appeared in year $t-1$ but not year t.

patent space for the first time, or that use words whose usage experiences fast growth across all patents. Hence, a patent using words such as "polypeptides", "clones", and "recombinant" would be classified as disruptive if its application year is 1995, but not in 2005. Symmetrically, patents relying on words whose usage decreases across all patents (i.e., using relatively older vocabulary such as "cassette," or "tape" in 1995) have lower (and possibly negative) levels of disruptiveness.

C Technological Breadth and Similarities

We also develop text-based measures capturing the technological breadth of each patent, as well as their similarities with other patents of economically linked firms. We posit that these characteristics are also relevant in predicting startup exits.

To measure the technological breadth of a patent, we first identify words that are strongly associated with a specific technological field using the six broad technological fields (f) defined by the first digit of the NBER technical classification.[14] Specifically, we count how often a given word (in N_t) is used by patents classified into each field in each year, and keep the two fields with the highest usage of the given word. We define a word as "specialized" (and associated with a field f) if its use in its most popular field is more than 150% that of its second most popular field. Each word is thus classified into one of the six fields of specialization or it is deemed an "unspecialized" word. For instance, words such as "bluetooth" and "wifi" are in the "Computer and Communication" field, and "acid" and "solvent" are in the "Chemicals" field. Second, we define as $w_{j,t,f}$ the fraction of patent j's specialized words that are classified into each field f. By construction, each $w_{j,t,f}$ lies in the [0,1] interval, and they sum to one for each patent j. We then define technological breadth as:

$$\text{Tech Breadth}_{j,t} = 1 - \sum_{f=1}^{6} w_{j,t,f}^2. \tag{3}$$

This measure is one minus the technological concentration of the patent's vocabulary. Patents have higher technological breadth if they amalgamate vocabularies from different

[14] "Chemicals", "Computer and Communication", "Drugs and Medicine", Electricity", "Mechanics", and "Others".

specialized technological fields and for which a wide range of knowledge is needed to develop and understand the invention.[15] In contrast, patents with lower breadth use vocabulary that primarily concentrates on one specialized technological field.

Next, we define three measures of technological similarity by directly comparing the vocabulary of a given patent to that of patents assigned to three specific groups: lead innovators, private U.S. firms, and foreign firms. To do so, we rely on the concept of cosine similarity (see Sebastiani (2002)), which is defined as the scalar product between each patent j's normalized word distribution vector $V_{i,t}$ and a normalized word vector aggregating the vocabulary specific to a given group of patents.[16]

To capture the similarity of a given patent j with patents of "Lead Innovators" (henceforth LI), we define LIs annually as the ten U.S. public firms with the most patent applications. This set varies over time as the importance of sectors and firms changes. LIs include Microsoft and Intel in 2005, IBM and Motorola in 1995, General Electric and Dow Chemical in 1985, and General Electric, Bell Telephone, and General Motors in 1935. For each set of LIs in year t, we first identify the set of patents applied for by the LIs over the past three years (i.e., from year $t-2$ to t). The aggregate LI word vector in year t corresponds to the equally-weighted average of the resulting normalized patent vectors. We then compute the similarity of any given patent to those of the LIs as:

$$\text{LI Similarity}_{j,t} = \frac{V_{j,t}}{\|V_{j,t}\|} \cdot \frac{V_{LI,t}}{\|V_{LI,t}\|}. \tag{4}$$

Because the word vector $V_{LI,t}$ aggregates word usage across patents of lead innovators in the last three years, patents exhibiting higher levels of LI similarity contain technologies that are textually close to those of lead innovators. In contrast, patents with low levels of LI similarity use text that is unrelated to that used in the patents of lead innovators, and thus are more distant technologies.

We use similar methods to 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. Specif-

[15] An illustrative example of a high technological breadth patent is the satellite (patent #2835548 from 1957), which required both mechanical and electronic technologies among others.

[16] The result is bounded in [0,1] and values close to one indicate closer textual similarity.

ically, we form the aggregate private firm (foreign firm) word vectors $V_{P,t}$ ($V_{F,t}$) as the equally-weighted average of the normalized vectors V of patent applications by private (or foreign) firms in year t.[17] We then define the similarity between each patent j and the contemporaneous patent applications of all private U.S. firms as:

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

Analogously, the similarity between patent j and those of foreign firms is:

$$\text{Foreign Similarity}_{j,t} = \frac{V_{j,t}}{\|V_{j,t}\|} \cdot \frac{V_{F,t}}{\|V_{F,t}\|}. \tag{6}$$

These measures are high for patents whose vocabulary is technologically close to that of patents assigned to private U.S. firms or to foreign firms, respectively. As proximity to private and foreign firms indicates more contested markets for innovation, these variables allow us to examine competitive effects in exit choices.

D Descriptive Statistics, Comparisons, and Examples

Table II presents descriptive statistics for our new text-based technological characteristics as well as existing patent variables from the literature for the full sample of patents (1930-2010). All variables are defined in Appendix B. We first focus on patents' technological disruptiveness. Across all patents between 1930 and 2010, we note that empirical distribution of technological disruptiveness is highly skewed. The first row of Panel A indicates that the average disruptiveness of patents is 1.69, the median is 1.27, and the 75^{th} percentile is 2.34. The observed asymmetry indicates that while the vast majority of patents contain incremental inventions, a smaller set of patents appear to be highly disruptive.

[Insert Table II about here]

We corroborate this intuition by comparing patents' technological disruptiveness with

[17] Because these groups contain very large numbers of patents, we aggregate over just the single year t. We also note that when a patent j belongs to a private U.S. firm or a foreign firm, we exclude it from the set of patents used to compute $V_{P,t}$ and $V_{F,t}$, respectively.

two variables commonly used to describe patent quality and economic value. First, we collect all citations for each patent as of early 2014 from Google Patents.[18] Second, for the set of patents assigned to public firms, we use their estimated value as measured by Kogan, Papanikolaou, Seru, and Stoffman (2016) using stock return changes around patent grants' announcements (henceforth KPSS). Panel B reveals positive and significant correlations, indicating that patent applications exhibiting higher levels of technological disruptiveness attract more ex post citations, and have larger economic values when granted. These positive relationships remain highly significant in regressions with fixed effects for application year, grant year, and technology category (and clustering standard errors by grant years).

Although a patent's technological disruptiveness is significantly related to future citations and economic value, the reported correlations range between 0.03 and 0.10, suggesting important differences across these measures, and especially between disruptiveness and future citations.[19] To better understand this difference, we study in detail the 25 patent applications occurring between 1980 and 2010 that have attracted the largest number of citations in the first five years following their grant. We display in Panel A of Table III the percentile of each such patent in the distribution of technological disruptiveness. Although citations are clear indications that patents are widely used ex-post, technological disruptiveness is specifically designed to capture new and fast-growing ideas ex-ante. This is evident in Panel A, where patents with high levels of disruptiveness tend to build-out novel ideas (e.g., the use of HTML in the mid-nineties), rather than refinements or synthesis of existing ideas (e.g., the iPhone or semiconductor advances in the mid-2000s). Indeed, the average percentile of technological disruptiveness for these 25 patents is 57, as the most disruptive patents refer to more nascent and viral technologies.

[Insert Table III about here]

Panel B displays similar figures for a collection of twelve unambiguous breakthrough

[18]We eliminate patents granted in early 2014 to avoid having a partial year of data.

[19]Unlike citations, our measure of technological fluidity is less exposed to truncation (e.g. Lerner and Seru (2017)) and it uses ex-ante information. Additionally, our measure can be computed for all patents, whereas economic value can only be computed for public firm patents (which represent only 28% of all patents).

patents, as identified by the USPTO.[20] The average technological disruptiveness of these patents is very high (84th percentile). The most disruptive patents in this set are "Complex computer" in 1944 (#2668661) and DNA modifications in 1980 (#4399216), both of which virtually created new industries. Other key inventions, such as the satellite, laser, and PageRank, use vocabulary that is new and rapidly growing around the time of their application. Interestingly, some of these breakthrough inventions are barely cited. For instance, the patents related to the invention of the "television" (#1773980) and the "helicopter" (#1848389) are in the lowest percentile of the cohort-adjusted distribution of citations. Yet, our new measure classifies these patents as highly disruptive.

Tables II and III also provide statistics for the other text-based measures. Unlike patents' technological disruptiveness, patent breadth is evenly centered around its average value of 0.42, indicating less skew specializations across patents. Remarkably, Table III indicates that patents' technological breadth and their originality (as defined by Trajtenberg, Henderson, and Jaffe (1997)) are largely unrelated.[21] For instance, software-related patents (e.g., #6964374, #7630986, or #7356679) display low levels of technological breadth, as they rely on a relatively narrow vocabularies. Yet, they rank high in terms of their originality. In contrast, breakthrough inventions display above-median levels of technological breadth. For example, the invention of the satellite, computer, and the jet engine combine broad scientific language.

Finally, we 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. Panel C of Table II further indicates that the text-based measures capture distinct dimensions of the technological nature of patents. Patents with higher

[20] Listed patents applied before 1960 come from a list of historical patents at http://www.uspat.com/historical/. More recent patents are noted for the revenue they generated.

[21] The construction of our complexity measure is somewhat similar, but we measure concentration based on the assignment of *words* to technology areas rather than citations. This has several advantages. First, our measure of complexity is well defined even for patents with zero or one backward citations. Second, technology links revealed by vocabularies are not influenced by strategic avoidance of citations and do not rely on patent examiners having a complete knowledge of the patent space. Third, our measure of complexity can pick up a reliance on a technology area even if no specific citation to that area is given. For example, a patent might use "textbook" information about chemistry to describe a portion of the invention without needing to cite a "Chemical" patent.

levels of disruptiveness tend to display lower breadth. They are also located nearer to patents of lead innovators than to patents assigned to private or foreign firms. We also note that patents nearer to lead innovators tend have lower technological breadth, while those most similar to private and foreign firms display higher breadth.

III The Decline in Technological Disruptiveness

In this section, we document the aggregate time-series properties of our text-based technological characteristics between 1930 and 2010. We then contrast these technological changes with aggregate changes in IPO and acquisition activities occurring between 1980 and 2010.

A Technological Changes in the Last Century

To track the evolution of technological disruptiveness in time series, we compute the aggregate stock of *Tech Disruptiveness*.[22] Figure II displays the time series from 1930 and 2010, smoothed using a four-quarters moving average. Several interesting patterns emerge, suggesting important changes were occurring in the technological characteristics of U.S. innovation. Although there is considerable fluctuation, the time series neatly maps the history of U.S. innovation. We easily detect periods of sharply increasing disruptiveness. The first peak occurs around 1950 with an average level of disruptiveness that is almost double the level of 1930. The period around 1950 is often considered a time of radical innovation in manufacturing technologies, featuring the invention of the television, transistor, jet engine, nylon, and xerography. A second peak occurs in the mid-seventies, corresponding to innovation related to the computer. The last two peaks of technological disruption appear in the late eighties and mid-nineties, reflecting waves of inventions related to genetics (e.g., methods of recombination) and the mass adoption of the Internet.

[Insert Figure II about here]

[22]To compute the aggregate stock of any patent variable, we first compute the sum of *Tech Disruptiveness* for patents applied for in a given quarter. We then compute a rolling depreciated sum of the prior 20 quarters, using a 5% quarterly rate of depreciation. Finally, we normalize the rolling stock by the number of patents applied for in the 20 prior quarters.

Despite these periodic surges in disruption, the 1930-2010 period is characterized by a protracted and steady long-term decline in the disruptiveness of U.S. patents. Between 1950 and 2010, the average level of technological disruptiveness has significantly decreased, with levels in 2010 being roughly one quarter that of 1950. Importantly, this decline is not due to changes in the composition of patents (e.g., shifts across technology classes) as we continue to observe a similar trend after we account for broad technology and location fixed effects. Rather, the decline in measured technological disruptiveness indicates a widespread deceleration in vocabulary usage growth rates among U.S. patents. This trend echoes recent research highlighting the increasing difficulty to generate new innovative ideas (e.g., Jones (2009) and Bloom, Jones, Reenen, and Webb (2017)).

Figure II also reveals important changes in the technological breadth of U.S. patents. The overall level of breadth steadily increases between 1930 to 1970. Beginning in the mid-seventies, there is a twenty-year period of growth in overall patent breadth which reaches a peak in the mid-nineties that was 20% above the 1970 level. In the most recent years, however, there is a large decline in the breadth of U.S. patents, dropping by about 25% between the mid-nineties and 2010.

Finally, we find an inverse U-shaped pattern in patent similarities over the last century. All three measures steadily increase until the eighties, as the text in the average U.S. patent during this period became increasingly similar to patents assigned to private U.S. firms, foreign firms, and lead innovators. Beginning in the eighties, however, these trends reversed, leading to marked declines in the similarity measures. The recent period is thus characterized by patents becoming both more specialized (i.e., lower technological breadth) and more distinct across firms.

B Technology, IPO, and Acquisitions (1980-2010)

We next examine the time series properties of IPOs and acquisitions. We obtain data on 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 or public firms) coded as a merger, acquisition of majority interest, or acquisition of assets giving the acquirer a majority stake.

[Insert Figure III about here]

Panel A of Figure III plots the number of IPOs for each quarter between 1980 and 2010. The patterns are similar to those reported by Gao, Ritter, and Zhu (2013), Doidge, Karolyi, and Stulz (2017), and Ewens and Farre-Mensa (2018). Remarkably, the evolution of IPO activity rather closely maps the aggregate dynamics of technological disruptiveness during this thirty-year period. The number of IPOs drops around 1990, coinciding with a decline in disruptiveness that follows the earlier surge in genetic science in the mid-1980s. There were more IPOs as the nineties progressed, when technological disruptiveness experienced a very large increase. The decline in IPO intensity then began in the early 2000s, when the average technological disruptiveness of U.S. patents also started to plummet. In the aggregate, the intensity of new public listings is substantially higher at times where the average technological disruptiveness of U.S. patent applications is elevated.

Panel B of Figure III plots the evolution of the number of acquisitions, both in total and separately for private firms. The number of acquisitions has increased since 1980, with a strong acceleration in the mid-nineties. We note subsequent declines in acquisitions in the aftermath of the tech bubble and the financial crisis. Yet, the number of acquisitions remains significantly higher since the mid-nineties when compared to the 1980-1995 period. This suggests a relationship between the surge in aggregate acquisitions and the decline in technological disruptiveness of U.S. patents. Although the pattern for trade sales is less striking than that for IPOs, the patterns suggest that acquisitions tend to be high when overall technological disruptiveness is lower.

IV Technological Disruptiveness and Startups' Exits

To better understand the interplay between technological changes and exits, we explore the cross-sectional relationship between our text-based technological characteristics and the decision of private firms to exit by going public or through trade sales. Ideally, we would study the exit strategy of *all* private firms that are plausible candidates for IPOs or acquisitions. Because data limitations preclude this, we focus on a large sample of venture-backed private firms, for which we observe both their technological characteristics and their exit choices.

A Sample of VC-backed Startups

We obtain data on VC-backed U.S. firms from Thomson Reuters's VentureXpert (Kaplan, Stromberg, and Sensoy (2002)), which contains detailed information about private startups 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 our attention to VC-backed firms that are granted at least one patent during the sample period. To assign patents to VC-backed firms, we follow Bernstein, Giroud, and Townsend (2016) and develop a fuzzy matching algorithm that matches the names of firms in VentureXpert to patent assignees obtained from Google Patents (see Appendix C for details). The result is an unbalanced panel of firm-quarter observations.[23] A firm enters our sample in the quarter it is founded (based on founding dates in VentureXpert) and exits the sample when its final outcome (IPO, acquisition, or failure) is observed based on the resolve date in VentureXpert. Firms still active as of November 2017 remain unresolved. We exclude firms if their founding date is missing or if it is later than the resolve date. The sample begins in 1980 to guarantee reliable data on outcomes and ends in 2010 because this is when the link to publicly traded firms created by Kogan, Papanikolaou, Seru, and Stoffman (2016) ends. Our final sample contains 561,982 firm-quarter observations, corresponding to 13,679 unique firms and 506,096 patent applications.

We obtain the technological characteristics for each firm-quarter by aggregating each patent-level variable (text-based and others) using their depreciated sums over the past 20 quarters using a quarterly depreciation rate of 5%. For example, the technological disruptiveness of firm i in quarter q corresponds to the depreciated sum of the disruptiveness

[23] 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 VC-backed startups are small and are unlikely to have complex corporate structures.

of all its patent applications in the past five years, normalized by the number of patents i applied for over that period.[24] We define the exit variables (IPO or sell-out) as binary variables equal to one if firm i experiences a given exit in quarter q. The construction of all variables is explained in detail in Appendix B.

[Insert Table IV about here]

As we found for the overall statistics reported in Table II, we find in Table IV a substantial asymmetry in the distribution of technological disruptiveness among VC-backed firms despite the aggregation of their patents. The other variables are overall in line with their aggregate counterparts, indicating that the technological characteristics of VC-backed firms are roughly representative of those in the economy at large. Table IV further indicates that the quarterly IPO rate (i.e. the number of IPOs in a quarter divided by the number of active firms in that quarter) is 0.38%, and the quarterly sell-out rate is 0.54%.[25]

[Insert Figure IV and Figure V about here]

Figure IV plots the time series of technological characteristics for all unresolved VC-backed firms in each quarter. For the sake of comparison, we compute the aggregate stock of each variable for the set of patents granted to VC-backed firms as in Figure II. The trends closely map those of the aggregate dynamics presented in Figure II, indicating that the technological changes occurring among VC-backed firms is mirroring economy-wide changes. In Figure V, we compare the evolution of IPO and sell-out rates for VC-backed firms to the aggregate patterns. We scale the aggregate patterns, i.e. the quarterly number of IPOs and acquisitions, by lagged real GDP to obtain exit rates. The upper figure shows that the evolution of IPO rates for VC-backed firms between 1980 and 2010

[24] Because *Foreign Similarity* and *LI Similarity* are non-trivially correlated (60% and 45%) with *Private Similarity*, we orthogonalize *Foreign Similarity* and *LI Similarity* by subtracting *Private Similarity*.

[25] We report additional information about the sample firms in Appendix D in Table A1. Relative to the founding date, IPOs and acquisitions play out over time. Of these, IPOs occur fastest on average, while failure (when explicitly listed by SDC) takes the longest. The average firm applies for its first patent after 4.78 years, and receives its first round of VC funding 6.64 years after its founding. All of these numbers are mechanistically shorter when measured relative to the first patent instead of the founding year.

closely follows aggregate IPO rates, especially the decreased IPO rates after 2000. The lower figure also shows agreement between the sell-out rates observed for VC-backed firms and the aggregate trend. In particular, we observe growth in acquisition activity in the latter part of the sample.

Although our sample of VC-backed firms does not include all firms that have the potential to go public or get acquired, our sample of VC-backed firms nevertheless represents a useful laboratory to study the interplay between technological changes and the evolution of IPOs and acquisitions. First, these firms account for a large share of the IPO market (Ritter (2017)) and the production of innovation (Gornall and Strebulaev (2015)). Second, their IPO and acquisition rates from the last thirty years appear comparable to the economy-wide patterns, as shown in the figures above.

B Validation

Our analysis rests on the ability of our measure of disruptiveness to identify startups that can create new markets or disrupt existing ones. To corroborate this interpretation, we examine whether the startups that we classify as technologically disruptive (based on their patents) do in fact operate in product markets that display higher risk of disruption.

First, for each startup, we identify public firms offering similar products and services, following an approach similar to Hoberg and Phillips (2016). We obtain product descriptions of startups from VentureXpert as reported in the year of their first round of funding. We then compute the (cosine) similarity between the text in each startup's product description (in year t) and that of all public firms, obtained from the product description section of their 10-K report (also in year t). Since 10-Ks became available in electronic format in 1997, we focus on 8,771 startups whose first funding round occurred after 1996. To identify the public firms operating in a given startup's market (in year t), we identify the 25 public firms that have the most similar product descriptions relative to the startup. This is done by measuring the cosine similarity between the startup's business description and public firm 10-K business descriptions and taking the 25 public firms with the highest cosine similarity.

Next, for each startup, we measure whether its 25 public firm peers are actually at risk of disruption. This is done by computing the fraction of paragraphs in each public firm's 10-K that mentions words related to technology-based disruption. In particular, we consider three measures. First, we search for paragraphs that contain words having the root "technol" and also a word having the root "change". Intuitively, these public firms peers are discussing exposures to technological change, a direct form of market disruption. Second, we identify paragraphs having the words with the root "technol" and also the word root "compet". Such firms are explicitly discussing competition on the margin of technological expertise. Third, we consider the more strict set of paragraphs containing at least one word with the roots "disrupt", "technol", and "compet". In order to be counted as a hit, a paragraph must contain all three. Table V reveals a positive association between the technological disruptiveness of startups (using our text-based patent measure) and these ex-post mentions of market disruption by related public firms. This strong positive relationship holds across specifications that include either year fixed effects or a more complete set set of fixed effects including year, technology, location, firm age, and firm cohort fixed effects.

[Insert Table V about here]

In a separate set of validation tests, we examine the product market characteristics of 1,579 startups in our sample that go public after 1997. Because these startups gain public status when they IPO, we are able to link their ex-ante patent-based technological disruptiveness measure to product market attributes that are only measurable when firms are publicly traded. For each IPO firm, we thus consider validation tests based on three characteristics of their post-IPO market (measured in the year of their initial listing): product market concentration (HHI), the total similarity to publicly traded peers (TSimm), and product market fluidity. These three measures are specific to each IPO firm and these tests are thus direct. This data is available since 1997 (see Hoberg and Phillips (2016) for competitiveness measures and Hoberg, Phillips, and Prabhala (2014) for fluidity measures). Our validation tests obtain from life cycle theory (Abernathy and Utterback (1978)), and we thus predict that disruptable markets are those for which superior new technologies are more likely to be discovered by competing early stage firms. For validation, we thus expect that IPO firms having higher values of our patent-based disruption measure should exit into more competitive, less differentiated and more fluid product markets.

[Insert Table VI about here]

Based on these direct firm-specific measures, Table VI confirms this ex post prediction. Newly-public firms with more technologically disruptive patents indeed exit into markets that are more fluid, contested, and thus disruptable. In contrast, IPO firms with less disruptive technology operate in more stable markets with less competition and more product differentiation. Put together, the results in Table V and Table VI confirm that our patent-based measure of technological disruptiveness is indeed strongly associated with ex-post predictions regarding product markets indeed being less stable and thus greater at risk of disruption.

C The Determinants of Startups' Exit Types

We now examine the relationship between startups' technological traits and their propensity to exit via IPO or sell-out.[26] Our central hypothesis is that firms with more disruptive technologies will be more likely to seek a public listing and less likely to sell out.

Our baseline specification relies on the competing risks regression approach of Fine and Gray (1999) that explicitly models the "risk" of choosing a particular exit in quarter q given that the firm is still unresolved at that time.[27] Firms enter the sample (i.e., become at risk of exiting) when they are founded. Their exit is modeled using competing hazards to reflect multiple potential exit strategies that are mutually exclusive. This approach allows us to estimate the relationships between startups' text-based technological traits and the full set of potential exits. In addition to the five text-based technology variables,

[26] As mentioned in Section IV.A, VentureXpert frequently fails to code firms as failed when they stop operations. Given this data limitation, we do not directly report tests relating to failure outcomes.

[27] The use of a competing risk model is relatively rare in finance. One recent exception is Avdjiev, Bogdanova, Bolton, Jiang, and Kartasheva (2017), who examine the determinants of convertible capital choice by banks.

we also include controls for the log number of patent applications in the past five years and a dummy indicating firms with no applications in the past five years.

To ensure that our text-based variables are not capturing the effects of known patent variables from the existing literature, we further include the originality of startups' patents as well as their citations (both aggregated for each startup-quarter as is the case for our text-based variables). We also note that these existing citation measures are not ex-ante measurable, which is in contrast to our text-based measures (which we designed to avoid any possibility of look-ahead bias). Following the literature on IPOs and acquisitions, we also control for overall market activity using market relative valuation and stock returns as well as an identifier for the last quarter of the year (Lowry (2003) and Pastor and Veronesi (2005)). We cluster the standard errors by startup to account for any potential within-startup dependencies over time.

[Insert Table VII about here]

The first two columns of Table VII confirm the key role played by technological disruptiveness in startups' exit choices. In the first column, we observe a strong positive link between startups' technological disruptiveness and their likelihood of exiting through an IPO in the next quarter. The point estimate is 0.218 with a t-statistic of 11.47. This indicates that a one standard deviation increase in startups' technological disruptiveness is associated with a 21.8% increase in the quarterly rate of IPOs.[28] On the other hand, column (2) reveals that the odds of exiting via a trade sale are negatively related to technological disruptiveness. Indeed, the estimated coefficient is negative (-0.179) and statistically significant with a t-statistic of -7.86. Hence a one standard deviation increase in disruptiveness indicates a 17.9% decrease in the sell-out rate. In addition to being statistically significant, the baseline estimates reveal economically large relationships.

Table VII further reveals that the other text-based technological characteristics are also important determinants of startups' exits. We observe for instance that firms' tech-

[28] As explained in Fine and Gray (1999), regression coefficients from a sub-distribution hazard model denote the magnitude of the relative change in the sub-distribution hazard function associated with a one-unit change in the given covariate. Therefore, estimated coefficients reflect the relative change in the instantaneous rate of the occurrence of the event in those subjects who are event-free.

nological breadth is positively related to IPO incidence, and is negatively related to trade sales. Intuitively, high breadth technologies are difficult for other firms to emulate and are less redeployable toward other uses. As such, firms with more technological breadth appear less amenable to acquisitions and are more amenable to exit via IPOs. Table VII also indicates that firms whose patents are more similar to those of other private firms or foreign firms are significantly less likely to exit through sell-outs (t-statistic of -7.84) and are marginally more likely to go public (t-statistic of 1.73). These results are in line with the negative link between product market similarity and the likelihood of being a target documented in Hoberg and Phillips (2010) for public firms. When a firm has more peers, it is more easily replaced and hence the rate of acquisition is lower for these firms. In contrast, firms holding patents that are more similar to that of lead innovators are significantly more likely to sell out (t-statistic of 2.76). This result is consistent with the idea that startups cater their innovation to the need of lead innovators to increase their acquisition odds, as predicted by Phillips and Zhdanov (2013) and supported by Wang (2017).

The lower portion of Table VII shows that the control variables are also relevant. For example, firms are more likely to exit via IPO after periods of strong overall stock market performance, consistent with growth options or market timing stories. Periods characterized by higher overall valuations have more IPOs and sell-outs. We also find that future citations (originality) are positively (negatively) associated with exit via both IPO and trade sale. Importantly, the inclusion of these variables does not impact the estimate of our text-based characteristics, which confirms that our measures of technological disruptiveness and breadth are distinct from these existing variables.

In the last two columns of Table VII, we report estimates from linear probability models where the dependent variables are indicators for whether a given exit occurs in a given quarter. Although this approach ignores the potential dependence across exits (i.e., competing risks), linear models allow us to include a wider array of fixed effects. We include year, state, technology, age, and cohort fixed effects to identify the link between exits and technological disruptiveness among firms of the same age, those receiving first funding at the same time, those operating in the same year and state, and those innovating in the same technological fields.[29] We find that our conclusions are largely unaffected, indicating that the effect of technological disruptiveness on startups' exits is unique. We also estimate (but do not report for brevity) separate logistic models for each exit type that include year, state, and technology fixed effects. These tests produce similar results.

[Insert Table VIII about here]

In Table VIII, we use the same specifications in Table VII, but we further control for startups' financing, as previous research reports that the amount of VC funding (a proxy for startups' implied valuation) predicts startups' exits (Cumming and Macintosh (2003)). These tests are important because our interpretation of the link between technological disruptiveness and startup exits could be due to VCs providing more funding to startups with more disruptive patents.[30] Table VIII indicates that this is not the case. To account for the possible role of funding, we include startups' cumulative VC funding (from founding to quarter $q-1$) and a binary variable identifying whether startups received funding in the last five years. Across all specifications, we confirm that the financing variables are strong determinants of startups' exit, especially sell-outs. However, our main result for technological disruptiveness is fully robust, indicating that our findings cannot be explained by financing.

Table IX explores the dynamic links between startups' technological disruptiveness and exits by increasing the measurement window for identifying startup exits from one quarter to five years using increments of one year. We focus on linear specifications that include the full set of fixed effects as described above, and only report coefficients for the technology variables for brevity. Panel A indicates that the positive associations between startups' technological disruptiveness and IPO incidence remains strong at all horizons. In contrast, Panel B reveals that the negative relation between technological disruptiveness and the propensity to sell out is only present at short horizons and then fades after two

[29]Technology fixed effects are based on the most common NBER technology category used in a firm's patents (see Lerner and Seru (2017)).

[30]We confirm this intuition in Table A2 of the Appendix.

years. We further report in Appendix D that our conclusions are stable throughout our sample, as the results hold across sub-periods.[31]

[Insert Table IX about here]

Finally, we also consider whether technological traits are related to the propensity of a startup to remain private for longer periods. This analysis is motivated by the evidence in Gao, Ritter, and Zhu (2013) and Ewens and Farre-Mensa (2018) that many startups remain private longer. Panel C of Table IX presents results from regressions of startups' odds of remaining private at different horizons on their current technological characteristics. Results reveal that startups with high technological disruptiveness exit more quickly than less disruptive startups. We also find that startups with higher overall technological similarity to other startups are more likely to stay private longer, and those that are more similar to foreign firms exit more quickly. Startups with high levels of private firm similarity are in more competitive markets and need additional time to establish stable market positions, whereas firms facing foreign competition need to achieve greater scale more quickly to avoid failure. Overall, these findings complement those of Gao, Ritter, and Zhu (2013) and Ewens and Farre-Mensa (2018).

V Disappearing IPOs and Surging Sell-Outs

Our results thus far indicate that, between 1980 and 2010, the technological disruptiveness of startups is systemically related to their propensity to exit via IPO and sell-out. Since technological disruptiveness has declined since the late nineties (see Figure IV), we examine whether the recent shift from IPOs to trade sales is explained by startups' technological changes.

A Prediction Errors for Startups' Exit

We use methods from the disappearing dividends literature (e.g., Fama and French (2001) or Hoberg and Prabhala (2009)) to assess the extent to which our technological characteristics can explain the disappearing IPOs and surging sell-out anomalies. We thus compare

[31] Therefore, it is unlikely that our results are affected by truncation biases coming for the "unresolved" status of recent startups or patent applications not yet granted as pointed out by Lerner and Seru (2017).

actual IPO rates to those predicted by models estimated using an ex ante period. To assess the impact of our new variables, we estimate predicted rates with and without our text-based technological variables.

We proceed in two steps. First, we estimate linear probability models where the dependent variable is the incidence of IPO exits in each quarter over the initial period 1980-1995 (the "pre-period"). The independent variables include firm and technological characteristics as well as state and technology fixed effects. Second, we compute predicted values of IPO incidence for each startup in the 1996-2010 period (the "post-period") using the fitted coefficients from the pre-period and actual values of the independent variables in the post-period. We then average the predicted IPO incidence rates across all startups in each quarter and compare them to the actual average IPO rate. We repeat these steps for sell-outs to compare actual and predicted sell-out exit rates.

[Insert Table X about here]

Table X presents the results. Panel A indicates that using only the control variables and fixed effects (i.e., excluding the text-based technology traits) to estimate the first-stage regression in the pre-period yields an average predicted quarterly IPO incidence rate of 0.63% in the post-period. This predicted incidence is substantially higher than the actual IPO rate, which is of 0.36% per quarter in the post-period. The predicted IPO rate is thus 75% higher than the actual rate. This confirms that IPO rates in the post-period are "abnormally low" and we thus replicate the disappearing IPOs anomaly. When we include the text-based technological characteristics in the model, the predicted IPO incidence in the post-period declines to 0.57%, which is still higher than the actual incidence rate. Although a significant portion remains unexplained, our estimates imply that our text-based technological traits explain 22% of the disappearing IPOs anomaly. The rest of Panel A indicates that our estimates are not sensitive to the definition of the pre- and post-periods or to the lags between the dependent and independent variables.

Panel B reports parallel analysis for sell-out rates. A benchmark first-stage linear model that excludes our technology variables estimated in the pre-period yields an average predicted sell-out incidence of 0.54% per quarter in the post-period. Compared to the actual rate of 0.63% per quarter, the base model's prediction is 14% lower than the actual rate, suggesting that the prevalence of sell-outs in recent years is "abnormally" high. We thus replicate the surging trade sales anomaly. When we include our text-based technology variables in the model, the prediction gap almost entirely disappears, as we obtain a predicted sell-out rate of 0.63% per quarter. When we alter specifications in the remainder of Panel B, we observe that changes in startups' technological characteristics explain between 20% and 68% of the surging trade sales. We conclude that roughly 55% (the average across specifications) of the surge in trade sales is explained by changes in their technological characteristics.

B The Role of Product Market Stability

We further propose that the role of technologies in explaining changes in IPO and trade sale rates likely differs across industries with varying product maturity. For example, as industries mature, the best ideas become "picked over" and later breakthrough inventions obtain only with very high search costs (Jones (2009) and Bloom, Jones, Reenen, and Webb (2017)). As breakthrough inventions decline, we expect an overall trend towards less disruptive and more narrow patents. A consequence is that disruptive inventions should vanish at increasing rates as industries mature, and hence initial industry conditions become crucial in predicting the future IPO rate. We consider the product life cycle in Abernathy and Utterback (1978) and posit that maturing markets in transition or those reaching maturity (e.g., markets that have reached a dominant product design) are likely to experience the most extreme decline in IPO rates. In contrast, very young "fluid" markets are more likely to have high IPO rates that might even grow during the earliest years as competing disruptive product designs are invented and commercialized (a process requiring some initial "time to build").

We follow Hoberg, Phillips, and Prabhala (2014) and compute the degree of product market fluidity in each firm's product market from 1980 to 2010. For our VC-backed startups, we use the business description text that is available at the time of the first funding round ("first money"). We then proceed in two steps. First, we compute the aggregate change in product description vocabulary used by startups as the year-over-year change in the frequency of usage across all business descriptions. This quantity is computed separately for each word and the result is stored in an aggregate vector containing the set of word frequency changes for all words (this procedure is similar to that in Equation (1)). Second, for a given startup, we compute the frequency-weighted average of the aggregate change vector where the weights are the frequency of words used by the startup in its own business description (this calculation is similar to that in Equation (2)). The resulting variable is a product fluidity measure similar to the one used in Hoberg, Phillips, and Prabhala (2014), but defined over all VC-backed startups receiving their first money between 1980 and 2010.

We use these firm-specific product fluidity measures to divide observations into above and below median fluidity sub-samples. We determine median breakpoints separately in each year. These above and below median groups have "fluid" and "stable" product markets, respectively. To test our prediction that mature and stable product markets experience more rapid declines in disruptive patents, and hence more severe declines in IPO rates, we separately assess the ability of our variables to explain the dearth of IPOs and the surge in trade sales in each sub-sample.

[Insert Table XI about here]

We present the results in Table XI. Consistent with predictions, we observe in Panel A that startups in stable markets are less likely to exit via IPO relative to startups in fluid markets. The actual average IPO rates are 0.32% and 0.40% per quarter for stable and fluid markets in the post-period (1996-2010), respectively. Moreover, we find that changes in startups' technological characteristics explain roughly 60% of the dearth of IPOs in stable markets. This figure ranges between 43% and 69% across different specifications. This finding illustrates a much stronger ability of technological traits to explain disappearing IPOs in transitional and more mature markets, as the full sample result was only 20%. In contrast, when we consider the fluid markets subsample, we find that changes in startups' technological attributes explain just 10% of the dearth of IPOs.

This figure ranges between 5% and 16% across specifications. Although market fluidity is critical for understanding IPO rates, Panel B indicates that fluidity has little effect in moderating our ability to explain the surging trade sales anomaly. We conclude that we explain roughly 55% of surging trade sales overall as explained above, and this result is not sensitive across sub-samples.

The results in this section suggest that the future IPO rate will likely depend critically on the extent to which new fluid industries emerge (such as those based on artificial intelligence, fintech, and other new technologies) and existing industries stabilize. Established industries should experience persistently declining IPO rates. Although more work is needed to draw policy recommendations, our findings suggest that these trends might be particularly unresponsive to policy changes, with subsidies to foster entrepreneurial activity being perhaps a possible exception.

VI Conclusions

We develop new measures of technological disruptiveness and other characteristics using textual analysis of 6,850,075 U.S. patents that were applied for between 1930 and 2010. We first document that these characteristics are highly influential in predicting which startups will exit via IPO or sell-out. These results are economically large and remain important after controlling for a host of known explanatory variables.

We find that startups with more disruptive patents are more likely to exit via IPO. Understanding the economics of disruptive patents is more intuitive when juxtaposed against patents that primarily refine or extend existing technologies. These technologies are also valuable, but they have less potential to establish independent markets for their owners. Intuitively, we find that startups having technologies with higher text-based patent similarity to firms with large patent portfolios are more likely to exit via sell-outs. These findings are consistent with the broader hypothesis that IPOs are favored by firms having technologies with the potential to establish independent market positions with strong defenses against rivals. Sell-outs are favored by firms whose technologies have less potential to create independent markets, and technologies that are more redeployable.

Such technologies are more valuable in the hands of existing firms with strong independent market positions.

We also find that technological traits change dramatically over our long sample period. For example, we document an economy-wide decline in technological disruptiveness that began after World War II. Because technological characteristics can explain exit choices at the micro level, we propose that wholesale changes in technological evolution might plausibly explain aggregate trends such as the recent decline in IPOs and the surge in sell-outs. By comparing actual IPO and sell-out intensities to predicted values that condition on our technological characteristics, we estimate that between 20% and 60% of the decline in IPOs is attributable to technological changes. Regarding sell-outs, roughly 55% of the surge can be explained by these same variables.

We note that our findings do not rule out other explanations for the decline in IPOs such as financing costs and regulatory change as they are not mutually exclusive. We believe that understanding the microfoundation driving the recent decline in IPOs can be valuable to academics, policy makers, and industry participants alike. Indeed, the appropriate policy response to declining IPOs depends critically on the specific economic forces that are behind the decline.

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, 2017, CoCo issuance and bank fragility, available at https://www.nber.org/papers/w23999.

Bates, Thomas, Kathleen Kahle, and Rene Stulz, 2009, Why do U.S. firms hold so much more cash than they used to?, *Journal of Finance* pp. 1985–2021.

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.

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

Bloom, Nicholas, Charles Jones, John Van Reenen, and Michael Webb, 2017, Are ideas getting harder to find?, Working Paper.

Chemmanur, Thomas, Jie He, 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.

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

Darby, Michael, and Lynne Zucker, 2018, Going public when you can in biotechnology, NBER Working Paper.

Doidge, Craig, Kathleen Kahle, Andrew Karolyi, and Rene Stulz, 2018, Eclipse of the public corporation or eclipse of the public markets?, Working Paper.

Doidge, Craig, Andrew Karolyi, and Rene Stulz, 2013, The U.S. left behind? Financial globalization and the rise of IPOs outside the U.S., *Journal of Financial Economics* pp. 546–573.

——— , 2017, The U.S. listing gap, *Journal of Financial Economics* pp. 464–487.

Ewens, Michael, and Joan Farre-Mensa, 2018, The deregulation of the private equity markets and the decline in IPOs, .

Fama, Eugene, and Kenneth French, 2001, Disappearing dividends: Changing firm characteristics or lower propensity to pay?, *Journal of Financial Economics* 60, 3–44.

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.

Gao, Xiaohui, Jay Ritter, and Zhongyan Zhu, 2013, Where have all the IPOs gone?, *Journal of Financial and Quantitative Analysis* 48, 1663–1692.

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

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.

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.

——— , and Nagpurnanand Prabhala, 2014, Product market threats, payout, and financial flexibility, *Journal of Finance* pp. 293–324.

Hoberg, Gerard, and Nagpurnanand Prabhala, 2009, Disappearing dividends, catering, and risk, *Review of Financial Studies* 22, 79–116.

Jones, Benjamin, 2009, The burden of knowledge and the death of the renaissance man: Is innovation getting harder?, *Review of Economic Studies* pp. 283–317.

Kaplan, Steven, 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, 2018, Measuring technological innovation over the long run, Working Paper.

Kogan, Leonid, Dimtris Papanikolaou, Amit Seru, and Noah Stoffman, 2016, Technological innovation, resource allocation and growth, *Quarterly Journal of Economics* forthcoming.

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

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

——— , Roni Michaely, and Ekaterina Volkova, 2017, Initial public offerings: A synthesis of the literature and directions for future research, Working Paper, Cornell University.

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 M., 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, 2017, Initial public offerings: Updated statistics, University of Florida.

——— , and Ivo Welch, 2002, A review of IPO activity, pricing, and allocations, *Journal of Finance* 57, 1795–1828.

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.

Wang, Xinxin, 2017, Catering innovation: Entrepreneurship and the acquisition market, Working Paper.

Figure I: Example of a Google Patent page

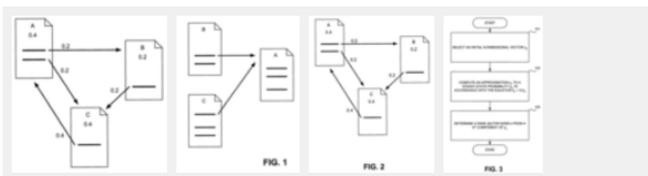
This figure shows the structure of a Google Patent page. The depicted patent is 6,285,999, commonly known as PageRank. Available at https://patents.google.com/patent/US6285999.

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.

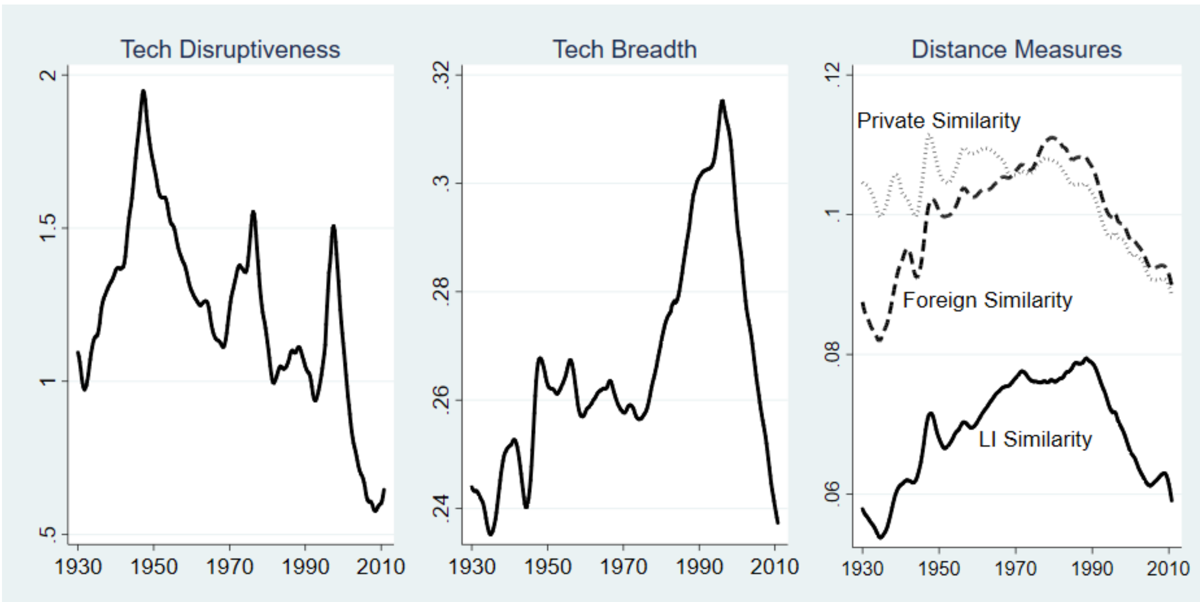
2. The method of claim **1**, wherein the assigning includes:

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: 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. To compute the aggregate stocks, we first compute the sum of each of the patent-level characteristics for patents applied for in a given quarter. We then compute a rolling depreciated sum of the prior 20 quarters, using a 5% quarterly rate of depreciation. Finally, we normalize the rolling stock by the number of patents applied for in the 20 prior quarters. The underlying patent level measures are winsorized at 1/99% level annually. The series presented are four quarter moving averages to smooth out seasonality.

Panel A: 1930-2010



Panel B: 1980-2010

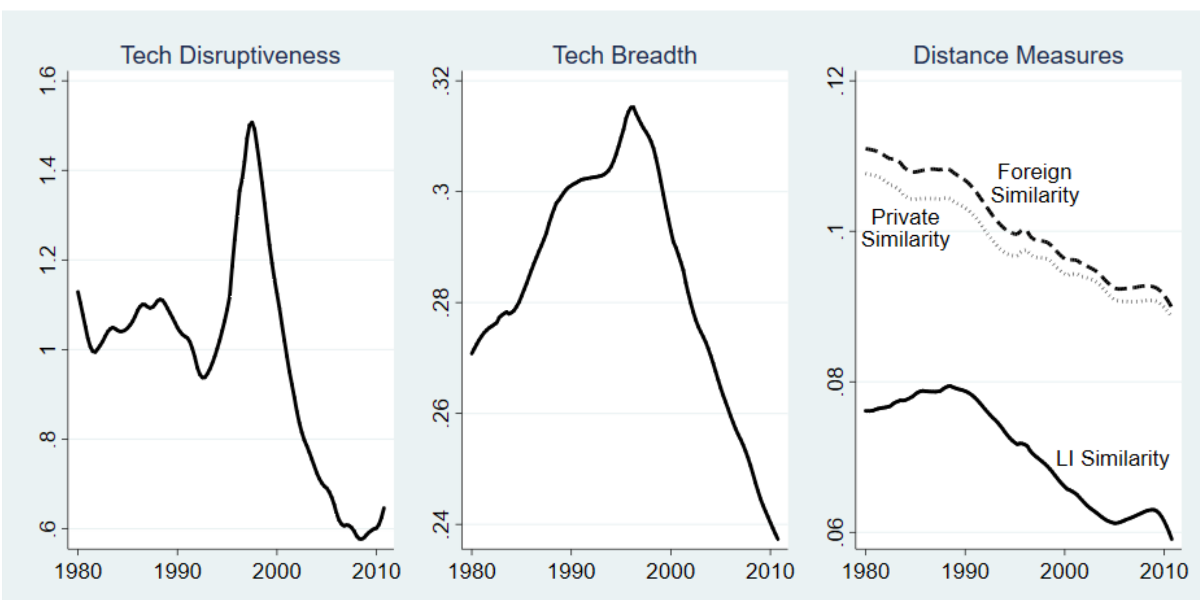
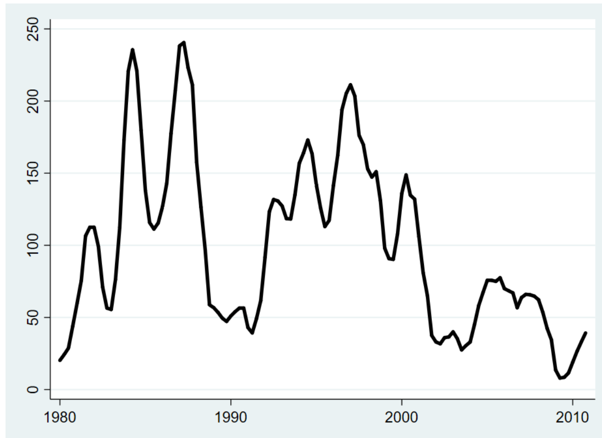


Figure III: Trends in Agregate IPOs and Acquisitions

This figure reports the time series of the number of IPOs in Panel A and acquisitions in Panel B. Data are from SDC. Panel B separately reports acquisition volume for all targets (dotted line) and private targets (solid line). The series presented are four quarter moving averages.

Panel A: IPO volume



Panel B: Acquisition volume

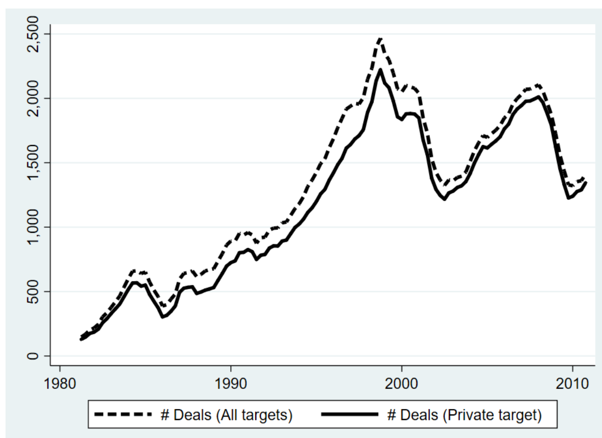


Figure IV: Trends in Technology Variables for VC-backed Startups

This figure reports characteristics of the aggregate corpus of patents held by VC-backed firms from 1980 to 2010. The variables are defined at the patent level in Section II. To compute the series below, we first compute the sum of each of the patent-level characteristics for patents applied for in a given quarter by VC-backed firms. We then compute a rolling depreciated sum of the prior 20 quarters, using a 5% quarterly rate of depreciation. Finally, we normalize the rolling stock by the number of patents applied for in the 20 prior quarters by VC-backed firms. The underlying patent level measures are winsorized at 1/99% level annually. The series presented are four quarter moving averages to smooth out seasonality.

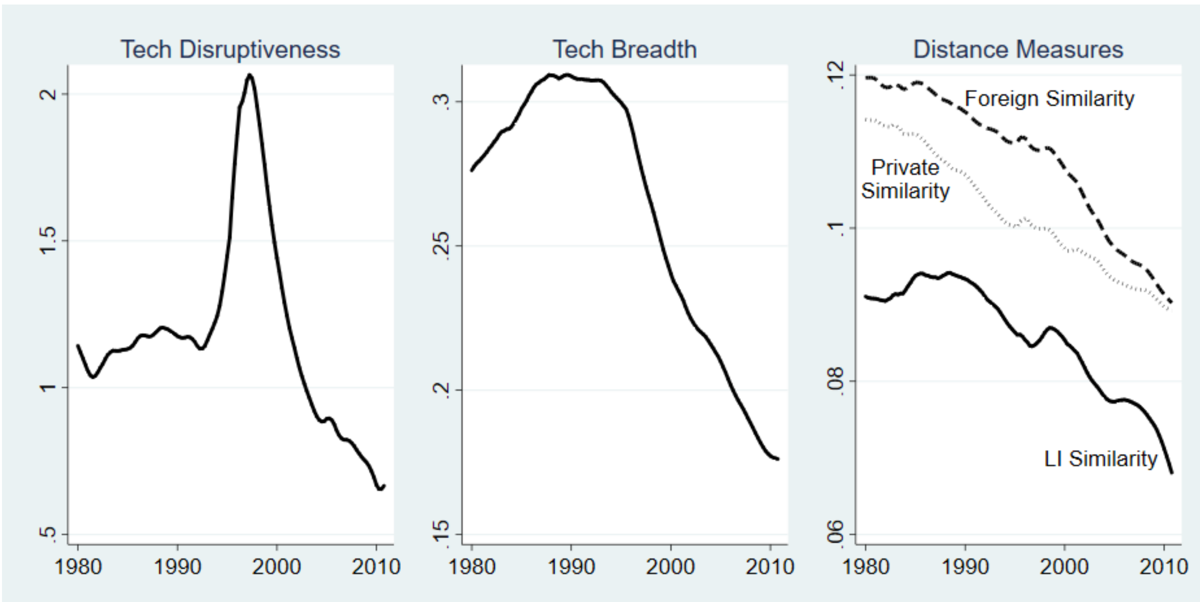


Figure V: Trends in IPOs and Sell-Outs of VC-backed Startups

This figure reports the propensity of firms to exit via IPO and acquisition. Aggregate trends are based on SDC data on IPOs and acquisitions of private targets and are reported in dashed lines as a fraction of lagged real GDP (left axis). Real GDP is in units of $100m. Trends in the VC-backed private firm sample are reported in solid lines as a percentage of firms that exist in the sample during the year (right axis). All series are reported as four quarter moving averages.

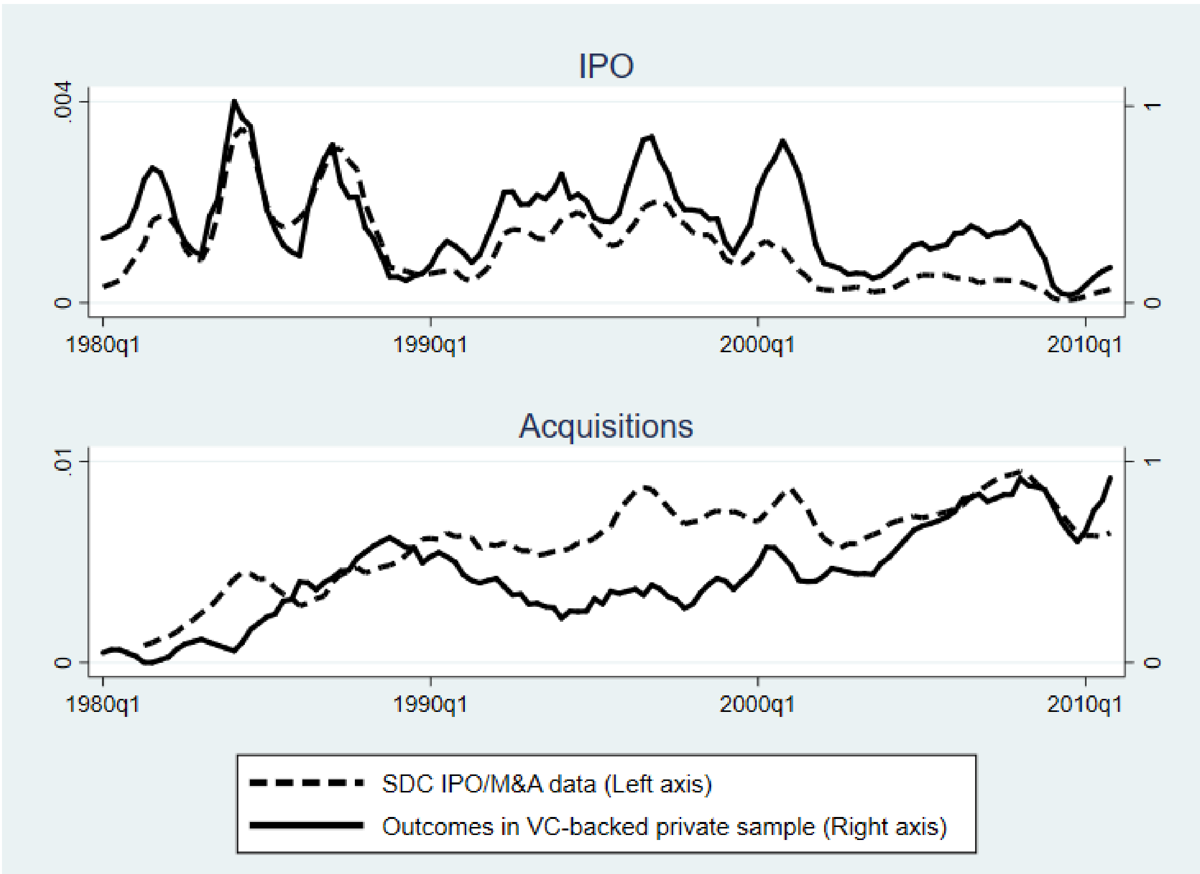


Table I: Changes in Patent Word Usage: Examples

This table reports, for five years throughout the sample period, how innovation has changed based on the text within patents. Panel A lists the ten words that have the largest year-over-year increase in use across all patents. Panel B lists the ten words that have the largest year-over-year decrease in use across all patents

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 descriptive statistics for patents granted between 1930 and 2010. The new technology variables *Tech Disruptiveness, Tech Breadth, Private Similarity, LI Similarity,* and *Foreign Similarity* are defined in Section II. Remaining variables are defined in Appendix B. Panel A reports summary statistics, where P25 and P75 denote the 25th and 75th percentiles. Panel B reports correlations between *Tech Disruptiveness* and preexisting measures of innovation. Panel C reports correlations between the new technology variables. All variables are winsorized at the 1/99% level annually.

Panel A: Summary statistics

	N	Mean	SD	P25	Median	P75
Tech Disruptiveness	6,594,248	1.69	2.25	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
KPSS Value	1,781,386	10.43	32.20	0.73	3.25	9.16
# of Cites	6,595,226	1.58	2.91	0.00	1.00	2.00
Originality	5,335,987	0.40	0.33	0.00	0.46	0.67

Panel B: Correlation with Innovation measures

		(1)	(2)	(3)
(1)	Tech Disruptiveness	1.00		
(2)	Log(1+KPSS Value)	0.06***	1.00	
(3)	Log(1+Cites)	0.03***	0.10***	1.00

Panel C: Correlation among text-based technology variables

		(1)	(2)	(3)	(4)	(5)
(1)	Tech Disruptiveness	1.00				
(2)	Tech Breadth	-0.19***	1.00			
(3)	Private Similarity	-0.10***	0.28***	1.00		
(4)	LI Similarity	0.03***	-0.09***	0.45***	1.00	
(5)	Foreign Similarity	-0.08***	0.25***	0.60***	0.74***	1.00

Table III: Technological Disruptiveness: Examples of Important Patents

For each patent, we report the percentile of three measures of innovation and two measures of complexity. The innovation measures we report are *Tech Disruptiveness* ("Disrpt"), *Cites* (received within five years of the patent grant), and *KPSS Value* ("KPSS"). The complexity measures we report are *Tech Breadth* ("Brdth") and *Originality* ("Orig"). The column labeled "Diff" reports the difference between the percentile of *Tech Breadth* and *Originality.* Red cells indicate higher values within a set of measures and blue cells indicate lower values within a set of measures.

		Innovation measures			Complexity measures			
Patent	Year	Dsrpt	Cites	KPSS	Brdth	Orig	Diff	Note
Panel A: Top 25 patents from 1980-2010 by citations								
7,674,650	2006	17	100	—	40	55	-15	Semiconductor/transistor
7,732,819	2008	14	100	—	42	55	-13	Semiconductor/transistor
7,501,293	2003	83	100	—	42	95	-53	Semiconductor/transistor
7,468,304	2006	13	100	18	26	1	25	Semiconductor/transistor
7,663,607	2004	67	100	96	29	80	-51	Multipoint touchscreen (Apple)
5,572,643	1995	99	100	—	15	72	-57	Early HTML use
7,462,862	2005	5	100	76	53	42	11	Semiconductor/transistor
7,453,065	2005	38	100	15	39	69	-30	Image sensor
7,453,087	2006	31	100	15	39	55	-16	Semiconductor/transistor
6,964,374	1998	65	100	77	12	88	-76	Storing and accessing metadata
7,630,986	2000	94	100	—	15	88	-73	Secure data interchange
7,632,985	2006	87	100	91	34	55	-21	Soybean biotech
7,411,209	2007	8	100	21	40	—	—	Semiconductor/transistor
7,181,438	2000	93	100	—	15	79	-64	User database system
7,402,506	2005	6	100	40	61	1	60	Semiconductor/transistor
7,356,679	2004	76	100	94	7	95	-88	Disk image capture
5,530,852	1994	100	100	93	16	1	15	Early HTML use
5,742,905	1994	99	100	—	12	52	-40	Personal communications
5,774,660	1996	98	100	—	20	86	-66	Early server management
7,479,949	2008	19	100	94	11	43	-32	Key iPhone patent
7,385,224	2005	10	100	14	25	81	-56	Semiconductor/transistor
5,608,786	1995	99	100	—	15	51	-36	Electronic comminucations
5,862,325	1996	99	100	—	9	84	-75	Electronic comminucations
5,708,780	1995	99	100	65	19	60	-41	Early server management
7,389,268	2002	7	100	—	12	23	-11	Tools for electronic trading
Panel B: Breakthrough patents								
1,773,980	1927	86	1	—	61	—	—	TV
1,848,389	1929	70	1	—	27	—	—	Helicopter
2,404,334	1941	73	95	—	77	—	—	Jet Engine
2,524,035	1948	82	100	71	46	87	-41	Transistor
2,569,347	1948	83	100	55	49	67	-18	Junction Transistor
2,668,661	1944	100	78	59	82	73	9	Modern digital computer
2,835,548	1957	81	68	—	99	—	—	Satellite
2,929,922	1958	90	99	87	60	—	—	Laser
4,237,224	1979	95	100	—	67	—	—	Cohen/Boyer patent
4,399,216	1980	99	99	—	73	1	72	"Axel" patent
4,681,893	1986	67	100	62	34	43	-9	Lipitor patent
6,285,999	1998	87	100	—	15	84	-69	PageRank (Google)

Table IV: Summary Statistics: Startup-Quarter Sample

This table presents summary statistics for the quarterly sample (1980-2010) of venture backed private firms. Firms are in the sample from their founding date until the quarter of their final outcome. Note that some firms remain private at the end of the sample period. The sample is further detailed in Section IV.A. The new *firm-level* technology variables *Tech Disruptiveness, Tech Breadth, Private Similarity, LI Similarity,* and *Foreign Similarity* are stock variables defined in Section IV.A. Remaining variables are defined in Appendix B. All variables are winsorized at the 1/99% level annually.

	N	Mean	SD	P25	Median	P75
Tech Disruptiveness	561,982	0.57	1.05	0.00	0.00	0.83
Tech Breadth	561,982	0.13	0.18	0.00	0.00	0.26
Private Similarity	561,982	0.05	0.06	0.00	0.00	0.10
LI Similarity	561,982	0.04	0.05	0.00	0.00	0.07
Foreign Similarity	561,982	0.05	0.06	0.00	0.00	0.09
Log(1+Firm Age)	561,982	3.18	1.18	2.48	3.26	3.89
No PatApps[q-1,q-20]	561,982	0.51	0.50	0.00	1.00	1.00
Log(1+PatApps[q-1,q-20])	561,982	0.69	0.92	0.00	0.00	1.10
Log(MTB) (q-2)	561,982	0.15	0.07	0.11	0.15	0.19
MKT Return [q-2,q-1]	561,982	0.01	0.13	-0.08	0.02	0.08
Q4	561,982	0.25	0.43	0.00	0.00	0.00
Originality	561,982	0.14	0.20	0.00	0.00	0.28
Log(1+Cites)	561,982	0.43	0.64	0.00	0.00	0.79
IPO rate (x100)	561,982	0.38	6.15	0.00	0.00	0.00
Sell-Out rate (x100)	561,982	0.54	7.32	0.00	0.00	0.00

Table V: Validity Tests: Public Peer Discussion

This table presents validity tests based on textual analysis of the 10-Ks of public peers of startup firms. The sample is a cross-section of observations where startups receive their first round of funding. As discussed in Section IV.B, we link each startup to the 25 public firms whose product descriptions—reported by VenturXpert as of the first VC round—are most similar. In columns (1)-(3), the dependent variable is the average fraction of paragraphs in public peers 10-Ks that mention words with roots "technol" and "change". In columns (4)-(6), the dependent variable is based on paragraphs with a root of "technol" and "compet". In columns (7)-(9), the dependent variable is based on paragraphs with a root of "technol" and "compet", together with either "disrupt", "change", or "obsoles". Thus, these variables measure the intensity with which the public peers of a given startup discuss technology-based market disruption, as discussed in Section IV.B. All variables are defined in Appendix B and are winsorized at the 1/99% level annually. We include year fixed effects for the year of IPO. Adjusted R^2 is reported as a percentage. Standard errors are heteroskedastic robust and are reported in parentheses. The symbols ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Discussion of:	"Tech" and "Change"			"Tech" and "Competition"			"Tech", "Competition", "Disrupt"		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Tech Disruptiveness	0.010***	0.007***	0.005**	0.042***	0.034***	0.017***	0.007***	0.005***	0.003*
	(3.94)	(3.00)	(2.57)	(6.20)	(5.29)	(2.58)	(4.44)	(3.19)	(1.95)
Constant	0.360***			1.376***			0.244***		
	(138.06)			(185.59)			(136.96)		
Observations	8,771	8,771	8,643	8,771	8,771	8,643	8,771	8,771	8,643
R2 (%)	0.2	5.4	30.6	0.4	5.2	22.4	0.2	7.0	25.4
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Technology FE	No	No	Yes	No	No	Yes	No	No	Yes
Location FE	No	No	Yes	No	No	Yes	No	No	Yes
Firm Age FE	No	No	Yes	No	No	Yes	No	No	Yes
Firm Cohort FE	No	No	Yes	No	No	Yes	No	No	Yes

Table VI: Validity Tests: Post-IPO competition

This table presents validity tests based on a subsample of the 1,579 startups that go public after 1997 where we are able to merge in both public firm identifiers (GVKEY) and obtain data on the product space of the firm (on startups' IPO year). The dependent variables *HHI* and *TSimm*, from Hoberg and Phillips (2016), are text-based measures of industry concentration and total similarity among a firm's public peers, respectively. *Product Mkt Fluidity* is from Hoberg, Phillips, and Prabhala (2014). All variables are defined in Appendix B and are winsorized at the 1/99% level annually. We include year fixed effects for the year of IPO. Adjusted R^2 is reported as a percentage. Standard errors are heteroskedastic robust and are reported in parentheses. The symbols ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	HHI (1)	Log(TSimm) (2)	Product Mkt Fluidity (3)
Tech Disruptiveness	-0.009***	0.143***	0.404***
	(-3.06)	(6.60)	(5.10)
Observations	749	749	708
R2 (%)	1.3	5.2	9.0
Year FE	Yes	Yes	Yes

Table VII: The Determinants of Startups' Exits - Baseline

This table presents cross-sectional tests relating a firm's ex-ante technological traits and its ultimate outcome. The outcomes we consider are IPO and sell-out (acquisition). The sample is a quarterly panel of venture backed private firms from 1980-2010 and is described in Section IV.A. Columns (1)-(2) use a competing risk hazard model and columns (3)-(4) use an OLS linear probability model. To facilitate interpretation, coefficients for OLS estimates report the incremental % change in a given outcome. Independent variables are lagged one quarter unless explicitly noted and all controls are standardized for convenience, except for the *Q4* and *No PatApps[q-1,q-20]* dummy variables. All variables are defined in Appendix B. *Li Similarity* and *Foreign Similarity* are orthogonalized relative to *Private Similarity*. All variables are winsorized at the 1/99% level annually. Technology fixed effects are based on the most common NBER-technology category across a firm's patents. Location fixed effects are based on the state reported in VentureXpert, where we combine international firms into one category. Adjusted R^2 is reported as a percentage. Standard errors are clustered by firm and 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)
Tech Disruptiveness	0.218***	-0.179***	0.081***	-0.058***
	(11.47)	(-7.86)	(5.53)	(-4.16)
Tech Breadth	0.424***	-0.200***	0.103***	-0.123***
	(9.87)	(-5.74)	(4.81)	(-5.40)
Private Similarity	0.105*	-0.390***	0.024	-0.293***
	(1.73)	(-7.84)	(0.88)	(-8.95)
LI Similarity	0.097**	0.109***	0.008	-0.004
	(2.08)	(2.76)	(0.35)	(-0.13)
Foreign Similarity	0.060*	-0.113***	0.036**	0.011
	(1.93)	(-4.35)	(2.53)	(0.66)
No PatApps[q-1,q-20]	1.258***	-1.975***	0.333***	-1.170***
	(10.48)	(-23.14)	(6.36)	(-15.96)
Log(1+PatApps[q-1,q-20])	0.266***	0.029	0.162***	-0.014
	(10.59)	(1.31)	(8.58)	(-0.68)
Log(MTB) (q-2)	0.241***	0.131***	0.151***	0.030
	(11.76)	(8.03)	(5.51)	(0.86)
MKT Return [q-2,q-1]	0.251***	0.022	0.029***	0.028**
	(10.52)	(1.07)	(2.99)	(2.05)
Q4	-0.031	0.088*	0.187***	0.241***
	(-0.50)	(1.76)	(5.13)	(5.69)
Originality	-0.107***	-0.131***	-0.031**	-0.112***
	(-3.31)	(-5.32)	(-2.19)	(-7.00)
Log(1+Cites)	0.104***	0.163***	0.050***	0.118***
	(3.07)	(6.57)	(3.31)	(6.13)
Year 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	559,866	558,965	553,897	553,897
R2 (%)	N/A	N/A	0.4	0.6

Table VIII: Determinants of Startups' Exits - Financing

This table presents cross-sectional tests relating a firm's ex-ante technological traits and its ultimate outcome. Each of the models repeats the corresponding model from Table VII, but adds endogenous financing controls. *log(CumVCFunding)* is the log of cumulative VC funding the firm receives between its founding and $q-1$. *No Funding[q-1,q-20]* is a control equal to one if the firm has not received funding in the prior 20 quarters. For brevity, we only report the new financing controls and *Tech Disruptiveness*. To facilitate interpretation, coefficients for OLS estimates report the incremental % change in a given outcome, and *Tech Disruptiveness* is standardized. Independent variables are lagged one quarter unless explicitly noted. All variables are winsorized at the 1/99% level annually. Adjusted R^2 is reported as a percentage. Standard errors are clustered by firm and 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)
Tech Disruptiveness	0.224***	-0.163***	0.073***	-0.068***
	(11.95)	(-7.38)	(4.98)	(-4.88)
log(CumVCFunding)	0.031***	0.169***	0.033***	0.101***
	(3.20)	(15.40)	(7.07)	(20.83)
No Funding[q-1,q-20]	-1.029***	-2.578***	-0.349***	0.059
	(-9.79)	(-11.34)	(-9.21)	(1.61)
Controls	Yes	Yes	Yes	Yes
Year 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	559,866	558,965	553,897	553,897
R2 (%)	N/A	N/A	0.5	0.8

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

This table presents dynamic cross-sectional tests relating a firm's ex-ante technological traits and its ultimate outcome over several horizons. In Panel A, column 1 repeats the OLS model examining IPO exits from column 4 in Table VII. Columns 2-6 subsequently replace the one-period ahead IPO exit indicator with longer horizons. We repeat this analysis for sell-outs in Panel B. Panel C examines whether a firm is still private (i.e. no IPO, or sell-out). In all models, the sample, independent variables, and coefficient interpretation are the same as the OLS models in Table VII. Independent variables are lagged one quarter and standardized for convenience. *LI Similarity* and *Foreign Similarity* are orthogonalized relative to *Private Similarity*. For brevity, the control variables and fixed effects are omitted. Standard errors are clustered by firm and are reported in parentheses. The symbols ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Exit within next:	Qtr (1)	Year (2)	2 Years (3)	3 Years (4)	4 Years (5)	5 Years (6)
Panel A: Exit by IPO						
Tech Disruptiveness	0.081***	0.302***	0.473***	0.551***	0.619***	0.626***
	(5.53)	(5.53)	(4.91)	(4.30)	(4.07)	(3.65)
Tech Breadth	0.103***	0.403***	0.796***	1.099***	1.316***	1.456***
	(4.81)	(5.09)	(5.46)	(5.56)	(5.42)	(5.22)
Private Similarity	0.024	0.065	0.066	0.003	0.129	0.250
	(0.88)	(0.65)	(0.37)	(0.01)	(0.44)	(0.75)
LI Similarity	0.008	-0.001	-0.039	-0.161	-0.226	-0.275
	(0.35)	(-0.01)	(-0.24)	(-0.73)	(-0.83)	(-0.88)
Foreign Similarity	0.036**	0.165***	0.332***	0.490***	0.590***	0.635***
	(2.53)	(3.03)	(3.32)	(3.57)	(3.46)	(3.22)
Panel B: Exit by Sell-Out						
Tech Disruptiveness	-0.058***	-0.187***	-0.226**	-0.228	-0.221	-0.165
	(-4.16)	(-3.34)	(-2.04)	(-1.46)	(-1.13)	(-0.73)
Tech Breadth	-0.123***	-0.523***	-0.986***	-1.372***	-1.709***	-1.870***
	(-5.40)	(-5.85)	(-5.78)	(-5.61)	(-5.51)	(-5.14)
Private Similarity	-0.293***	-0.980***	-1.357***	-1.250***	-0.732*	0.078
	(-8.95)	(-7.66)	(-5.54)	(-3.61)	(-1.71)	(0.16)
LI Similarity	-0.004	0.046	0.325	0.744**	1.169***	1.780***
	(-0.13)	(0.43)	(1.58)	(2.51)	(3.14)	(4.10)
Foreign Similarity	0.011	-0.013	-0.216*	-0.564***	-0.873***	-1.295***
	(0.66)	(-0.20)	(-1.72)	(-3.12)	(-3.78)	(-4.79)
Panel C: Still Private						
Tech Disruptiveness	-0.068***	-0.286***	-0.576***	-0.751***	-0.839***	-0.835***
	(-3.16)	(-3.52)	(-3.83)	(-3.70)	(-3.48)	(-3.12)
Tech Breadth	-0.010	-0.042	-0.160	-0.288	-0.275	-0.255
	(-0.29)	(-0.32)	(-0.67)	(-0.89)	(-0.69)	(-0.57)
Private Similarity	0.244***	0.905***	1.452***	1.682***	1.435***	1.076*
	(5.36)	(5.29)	(4.62)	(3.93)	(2.78)	(1.84)
LI Similarity	0.018	0.068	-0.025	-0.235	-0.432	-0.782
	(0.48)	(0.47)	(-0.09)	(-0.64)	(-0.96)	(-1.55)
Foreign Similarity	-0.090***	-0.322***	-0.492***	-0.493**	-0.592**	-0.585*
	(-3.73)	(-3.47)	(-2.85)	(-2.09)	(-2.05)	(-1.79)

Table X: Prediction Errors for IPO and Sell-Out Rates

This table presents a comparison of the out-of-sample performance of predictive models with variables standard in the IPO literature (the "Base" model) and a model which augments the "Base" model with the new text-based variables (the "BFH" model). Panel A examines IPO exits and Panel B examines sell-outs. In a given test (column 1), we estimate a predictive OLS model for a look-ahead horizon listed in column 2 on the sample period beginning in 1980 and ending before the testing period listed in column 3. 1Q, 1Y, 2Y, and 3Y denote forward looking windows of 1 quarter, 1 year, 2 years, and 3 years, respectively. The cross-sectional sample and underlying data are described in Table VII. We exclude time-varying fixed effects but keep technology and location fixed effects. We then apply the coefficients from the trained models on the testing sample listed in column 3, and average these predictions for each quarter. The true average quarterly exit rates for the testing period are listed in column 4 and the model predicted exit rates from the "Base" and "BFH" models are listed in columns 5 and 6. Columns 7 and 8 list the average prediction errors for each model. Column 9 reports the percentage improvement for the "BFH" model relative to the "Base" model. All probabilities in columns (4)-(8) are reported as percentages.

	Look	Testing	True	Model Prob.		Error		BFH
Test	Ahead	Period	Prob.	Base	BFH	Base	BFH	Impr
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Panel A: IPO Exits				
1	1Q	[1996,2010]	0.36	0.63	0.57	-0.28	-0.22	**22%**
2	1Q	[1998,2010]	0.32	0.62	0.55	-0.30	-0.23	**24%**
3	1Q	[2000,2010]	0.29	0.63	0.57	-0.34	-0.28	**18%**
4	1Y	[1996,2010]	1.40	2.47	2.23	-1.07	-0.83	**22%**
5	2Y	[1996,2010]	2.75	4.98	4.48	-2.23	-1.73	**22%**
6	3Y	[1996,2010]	4.05	7.60	6.89	-3.55	-2.85	**19%**
				Panel B: Sell-Out Exits				
7	1Q	[1996,2010]	0.63	0.54	0.63	0.10	0.00	**98%**
8	1Q	[1998,2010]	0.68	0.48	0.59	0.20	0.09	**54%**
9	1Q	[2000,2010]	0.72	0.50	0.63	0.22	0.09	**57%**
10	1Y	[1996,2010]	2.59	2.18	2.46	0.41	0.13	**68%**
11	2Y	[1996,2010]	5.35	4.23	4.61	1.12	0.74	**34%**
12	3Y	[1996,2010]	8.19	6.12	6.53	2.06	1.66	**20%**

Table XI: Prediction Errors for IPO and Sell-Out Rates - Market Stability

This table presents a comparison of the out-of-sample (OOS) performance of two predictive models in two samples. We compare a "Base" model using variables standard in the IPO literature to a "BFH" model which augments the "Base" model with the new text-based variables. Panel A examines IPO exits and Panel B examines sell-outs. The procedure is analogous to that described in Table X, except each test is repeated for two subsamples: *Stable Markets* and *Fluid Markets*, which are defined in Section V.B. We omit the model-implied OOS probabilities to conserve space.

				Stable Markets				Fluid Markets			
Test (1)	Look Ahead (2)	Est. (3)	Testing Period (4)	True Prob. (5)	Base Error (6)	BFH Error (7)	BFH Impr (8)	True Prob. (9)	Base Error (10)	BFH Error (11)	BFH Impr (12)
				Panel A: IPO Exits							
1	1Q	OLS	[1996,2010]	0.32	-0.15	-0.07	**56%**	0.40	-0.43	-0.40	**6%**
2	1Q	OLS	[1998,2010]	0.31	-0.15	-0.05	**65%**	0.35	-0.48	-0.42	**12%**
3	1Q	OLS	[2000,2010]	0.31	-0.14	-0.04	**69%**	0.28	-0.57	-0.54	**5%**
4	1Y	OLS	[1996,2010]	1.27	-0.47	-0.18	**62%**	1.56	-1.79	-1.62	**9%**
5	2Y	OLS	[1996,2010]	2.50	-0.95	-0.46	**52%**	3.08	-3.71	-3.09	**17%**
6	3Y	OLS	[1996,2010]	3.68	-1.58	-0.89	**43%**	4.54	-5.82	-4.90	**16%**
				Panel B: Sell-Out Exits							
7	1Q	OLS	[1996,2010]	0.61	0.07	0.02	**73%**	0.66	0.11	-0.04	**61%**
8	1Q	OLS	[1998,2010]	0.65	0.17	0.11	**36%**	0.71	0.24	0.07	**68%**
9	1Q	OLS	[2000,2010]	0.68	0.18	0.11	**42%**	0.76	0.26	0.10	**63%**
10	1Y	OLS	[1996,2010]	2.48	0.31	0.22	**28%**	2.71	0.50	-0.08	**83%**
11	2Y	OLS	[1996,2010]	5.06	0.90	0.97	**-7%**	5.65	1.29	0.29	**78%**
12	3Y	OLS	[1996,2010]	7.66	1.69	1.91	**-13%**	8.74	2.33	0.98	**58%**

A Defining the Entity Type of Patents' Assignees

In order to explore how a VC-backed firm's outcomes are related to competitive pressures and potential buyers, we augment the patent level dataset to denote if a patent is granted to (A) a private, domestic U.S. firm, (B) an international firm, or (C) a U.S. public firm.

First, we find all patents assigned to public firms. We obtain the GVKEY for assignees from the NBER patent dataset, and augment this with 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 indicates 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 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 are able to 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".

B Variable Definitions

Patent level variables

Tech Disruptiveness	See Equation 2 and Section II.B.
Tech Breadth	See Equation 3 and Section II.C.
LI Similarity	See Equation 4 and Section II.C.
Private Similarity	See Equation 5 and Section II.C.
Foreign Similarity	See Equation 6 and 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 for to the NBER technology codes so that *Tech Breadth* and *Originality* are based on the same granularity of technology classifications.

Firm-quarter variables

Tech Disruptiveness	The depreciated sum of patent-level *Tech Disruptiveness* for patents the firm applied for over the prior 20 quarters. Quarterly depreciation is 5%. We normalize the depreciated sum by the number of patents the firm applied for. See Section IV.A for more.
Tech Breadth	Converted to firm-quarter like *Tech Disruptiveness*.
Private Similarity	Converted to firm-quarter like *Tech Disruptiveness*.
LI Similarity	Converted to firm-quarter like *Tech Disruptiveness*.
Foreign Similarity	Converted to firm-quarter like *Tech Disruptiveness*.
Log(1+Cites)	Log of the stock of citations. Citations for a firm-quarter is the sum of the *# of Cites* (patent-level variable defined above) for patents the firm applies for in the quarter. Note that this is forward-looking. The stock is computed using a quarterly depreciation of 5%.
Originality	Converted to firm-quarter like *Tech Disruptiveness*.
No PatApps[q-1,q-20]	Dummy variable equal to one if the firm has not applied for a patent (which was eventually granted) during the last 20 quarters.
Log(1+PatApps[q-1,q-20])	The log of 1 plus the number of patent applications (which were eventually granted) made by the firm in the last 20 quarters.
Log(1+Firm Age)	Firm age is defined as the number of years since its founding date according to VentureXpert.
IPO	One if the firm goes public in the quarter, zero before.
Sell-out	One if the firm is acquired in the quarter, zero before.

Quarterly variables

Log(MTB) (q-2)	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 [q-2,q-1]	We compute quarterly market returns from Ken French's daily factor file using geometric compounding.
Q4	Equal to one if $t-1$ is the fourth quarter (and t is the first quarter) of the year.

C Matching patents to VentureXpert

A critical issue is that our research framework requires knowledge of the outcomes for private firms that receive patents—Does the firm go public, get acquired or simply remain private? Thus, 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, its final resolution (as IPO, acquisition, or unresolved) and date of resolution, 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. However, no commercially available link table exists. Hence, with the assistance of a computer science graduate student, 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.[32] 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 standarization routine from Nada Wasi.[33] 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, or 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, or 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, or 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

[32] Firms can receive patents before VC funding.

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

name. For each remaining patent, we keep as candidate matches any pair with equivalent name stems and the highest bi-gram match above 75%.

(c) A random subset of suggested matches, in addition 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.[34] 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 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.

[34]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 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 for research). 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 19,324 firm sample.

D Additional Tables

1. Table A1 presents information on the timing of key life events for firms in the main analysis sample.
2. Table A2 presents subsample tests of the main OLS models on the determinants of firm exit from Table VII. The subsamples are based on the date of the observation.
3. Table A3 presents tests of regression of startups' financing on their technological characteristics.

Table A1: Years between keys events for ventured backed private firm

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

Panel A: Events after the firm's founding

		Years between the firm's founding and event				
Event	N (firms)	Mean	SD	P25	Median	P75
First patent	13,679	4.78	12.16	0.50	2.25	6.50
VC funding	13,679	6.64	12.31	0.75	2.50	6.75
IPO	2,496	10.50	10.97	4.75	7.75	12.25
Acquisition	4,082	11.76	11.64	6.00	8.75	13.00

Panel B: Events after the firm's first patent

		Years between the firm's first patent and event				
Event	N (firms)	Mean	SD	P25	Median	P75
VC funding	13,679	1.86	8.57	-1.75	0.25	3.75
IPO	2,496	3.70	8.67	-0.25	3.25	7.00
Acquisition	4,082	7.66	7.23	3.75	6.25	10.25

Table A2: Subsample analysis of private firm exit: Time

This table repeats the OLS cross-sectional tests in columns (4)-(5) from Table VII on two subsamples. The tests relate a firm's ex-ante technological traits and its ultimate outcome. We split the sample based on the observation date. Even numbered columns include observations before January 1, 1996 and odd numbered columns include observations on or after January 1, 1996. In all models, the definition of independent variables and interpretation of coefficients are the same as the OLS models in Table VII. Independent variables are lagged one quarter and standardized for convenience. Note that we standardize variables *within* the subsample of the test. *Li Similarity* and *Foreign Similarity* are orthogonalized relative to *Private Similarity.* For brevity, the control variables are omitted. All variables are winsorized at the 1/99% level annually. Technology fixed effects are based on the most common NBER-technology category across a firm's patents. Location fixed effects are based on the state reported in VentureXpert, where we combine international firms into one category. Adjusted R^2 is reported as a percentage. Standard errors are clustered by firm and are reported in parentheses. The symbols ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Exit Type:	IPO		Acquisition	
Observation before/after 1995:	Before (1)	After (2)	Before (3)	After (4)
Tech Disruptiveness	0.146***	0.080***	-0.117***	-0.043**
	(3.86)	(5.52)	(-5.22)	(-2.52)
Tech Breadth	0.078	0.028	-0.187***	-0.181***
	(1.24)	(1.33)	(-3.53)	(-6.71)
Private Similarity	0.042	-0.003	-0.033	-0.297***
	(0.49)	(-0.12)	(-0.46)	(-8.25)
LI Similarity	0.065	-0.042*	0.154***	0.018
	(0.95)	(-1.74)	(2.83)	(0.57)
Foreign Similarity	-0.003	0.053***	-0.111***	-0.022
	(-0.06)	(3.61)	(-2.92)	(-1.17)
Observations	169,550	384,347	169,550	384,347
Firms	5,528	11,877	5,528	11,877
R2 (%)	0.4	0.3	0.4	0.4
Year FE	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes
Firm Age FE	Yes	Yes	Yes	Yes
Firm Cohort FE	Yes	Yes	Yes	Yes

Table A3: The Determinants of Startups' VC Funding
This table presents OLS cross-sectional tests relating a firm's ex-ante technological traits and its VC financing. The outcomes we consider are the log of cumulative VC funding the firm receives between its founding and quarter q, and a binary variable equals one if a firm receives a new round of VC financing in quarter q. For brevity, we only report the coefficients on the text-based technology variables. To facilitate interpretation, coefficients for OLS estimates report the incremental % change in a given outcome, and *Tech Disruptiveness* is standardized. Independent variables are lagged one quarter. All variables are winsorized at the 1/99% level annually. Adjusted R^2 is reported as a percentage. Standard errors are clustered by firm and are reported in parentheses. The symbols ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Ind. Variable:	Cum.Funds	New round	Cum.Funds		New Round	
Sample:	Whole	Whole	pre-95	post-95	pre-95	post-95
	(1)	(2)	(3)	(4)	(5)	(6)
Tech Disruptiveness	0.625***	0.179***	1.029***	0.405***	0.283***	0.137***
	(8.28)	(7.02)	(7.02)	(4.72)	(6.00)	(4.76)
Tech Breadth	-0.211*	-0.376***	0.241	-0.552***	-0.115	-0.546***
	(-1.80)	(-8.51)	(0.88)	(-3.96)	(-1.25)	(-10.41)
Private Similarity	1.662***	0.373***	1.406***	1.638***	0.469***	0.291***
	(10.36)	(6.71)	(3.91)	(8.93)	(3.96)	(4.62)
LI Similarity	0.854***	0.516***	1.155***	0.544***	0.705***	0.348***
	(6.19)	(10.53)	(3.98)	(3.32)	(7.33)	(5.96)
Foreign Similarity	-0.391***	-0.316***	-0.259	-0.461***	-0.324***	-0.292***
	(-4.36)	(-9.81)	(-1.25)	(-4.49)	(-4.46)	(-8.03)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	561,981	553,897	171,385	390,596	169,550	384,347
Firms	13,679	13,645	5,551	11,955	5,528	11,877
R2 (%)	2.4	31.0	2.9	1.7	22.8	26.4