RESEARCH

Swiss Finance Institute Research Paper Series N°19-22

Technological Disruptive Potential and the Evolution of IPOs and Sell-Outs

Donald Bowen
Lehigh University

Laurent Frésard
Università della Svizzera italiana and Swiss Finance Institute

Gerard Hoberg
University of Southern California

Technological Disruptive Potential and the Evolution of IPOs and Sell-Outs

Donald Bowen, Laurent Frésard, and Gerard Hoberg*

November 12, 2019

ABSTRACT

We study the determinants of startup exit methods using novel measures of technological characteristics constructed from patent text. Our main result is that startups with more potential to disrupt technological areas are 25% more likely to exit via IPO and 19% less likely to sell-out. These results suggest that IPOs are favored by startups that can carve out independent market positions, avoiding the need to share gains with an acquirer. We document an economy-wide decline in disruptive potential between 1930 to 2010, which can explain roughly 20% of the recent decline in IPOs, and 50% of the surge in sell-outs.

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

JEL classification: G32, G34, G24

---

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

I Introduction

Since the late 1990s, the number of private firms exiting via initial public offerings (IPOs) in U.S. markets has sharply declined. At the same time, the number of exits via acquisitions (i.e., sell-outs) has soared. Successful firms are nowadays more likely to sell-out to other (public or private) companies than seek independent public listings, a phenomenon that has recently garnered considerable attention in the media and policy circles.[1] In this paper, we show that U.S. startups developing technologies with disruptive potential are more likely to exit through IPOs than selling out. The role of technology in explaining exit choices is especially prevalent for startups with the potential to disrupt established technological areas as opposed to those introducing new technologies. We further document that the average disruptive potential of startups has markedly decreased in recent years and estimate that changes in startups' technological traits can explain 20% of the recent decline in IPOs and 50% of the surge in sell-outs.

Our analysis of startups' exits builds on the idea that rational entrepreneurs (and their backers) choose the exit option that maximizes the value of their equity stake. Using the theory of the firm (Grossman and Hart (1986) and Hart and Moore (1990)) and the exit theory of Bayar and Chemmanur (2011) as a foundation, we propose that these exit-specific valuations depend on technological characteristics such as the potential to disrupt established technologies, or to complement existing inventions through synergies. The relative attractiveness of IPOs compared to sell-outs thus hinges on the interactions between potential buyers' technologies and those of the startup, as well as the allocation of payoffs between parties. We posit that technologies with disruptive versus synergistic potential differ notably along both dimensions, and therefore trigger different exits.

By design, startups developing disruptive technologies offer limited synergistic value to other parties because disruptive inventions tend to be substitutes that can replace existing technologies (Acemoglu, Akcigit, and Celik (2014)). In addition, the economic success

[1]Various observers in the media and policy circles worry that the decline in new public listings reflects an erosion in the ability of U.S. financial markets to spur economic growth. 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).

of startups with disruptive potential can be achieved while remaining independent, as it is not necessary to integrate or seek strategic assistance from potential acquirers (Bayar and Chemmanur (2011)). This independence avoids the need to share future payoffs with another party. Hence startups with technologies with high disruptive potential should favor exiting via a public listing. In contrast, exiting by selling out should be favored by startups with technologies offering significant synergies to potential buyers. Acquisitions of synergistic technologies can improve existing processes or products, and buyers' resources can furthermore reduce financial constraints and foster product market success (Bena and Li (2014)). A sell-out is optimal when the complementary benefits of synergies overcome the cost of sharing future payoffs.

Testing our hypotheses requires the ability to measure technological characteristics at the patent and startup level. We exploit the voluminous text of all patents filed with the U.S. Patent and Trademark Office (USPTO) between 1930 and 2010 (6.6 million patents). We define a patent's disruptive potential as its potential to change the path of technological evolution and eventually disrupt established technologies or create new ones (Dahlin and Behrens (2005)).[2] We measure a patent's "technological disruptive potential" (henceforth "disruptive potential") based on the extent to which its vocabulary is new or growing fast across all contemporaneous patent applications. For example, the use of genetics words such as "peptide", "clone", or "recombinant" soared in 1995, reflecting concurrent breakthroughs in genome sequencing. Our measure would classify patents using such words in 1995 specifically as having high disruptive potential.

Because our goal is to estimate predictive models of startup exit, we measure disruptive potential using only ex ante measurable data from the text in all past and contemporaneous patents (see Dahlin and Behrens (2005) for a discussion of the distinction between ex ante and ex post innovation measures). The ex ante feature of our measure also eliminates look-ahead bias, reduces truncation bias (Lerner and Seru (2017)), and increases the measure's utility to practitioners (to predict outcomes and investor returns) and regulators

[2]We strictly follow the dictionary definition of "disruption", defined as "*a break or interruption in the normal course or continuation of some activity or process*"(https://www.merriam-webster.com/dictionary/disruption) or equivalently as "*an interruption in the usual way that a system, process, or event works*" (https://dictionary.cambridge.org/dictionary/english/disruption).

(to assess the impact of policies in real time).

We validate our measure of patent disruptive potential in four ways. First, we examine whether patents with higher ex ante disruptive potential change the path of technological evolution. We find that these patents receive significantly more citations and they simultaneously trigger reductions in the citation trajectory of preexisting and related patents. This latter test is based on a measure by Funk and Owen-Smith (2016), which is designed to identify structural "breaks" in the evolution of technology. Second, disruptive patents have higher economic value (estimated using stock returns as in Kogan, Papanikolaou, Seru, and Stoffman (2016)), suggesting that the market recognizes their potential. Third, we focus on historical breakthrough patents recognized by USPTO between 1930 and 2010 (including the television, computer, helicopter, and advances in modern genetics) and find that these patents have significantly higher ex ante disruptive potential. Fourth, we use textual similarity analysis to identify the publicly traded industry rivals for each startup. As direct validation, these public firms discuss market disruption more in their 10-Ks when they operate alongside startups with disruptive potential.

Our primary analysis focuses on the exit decisions of 9,167 VC-backed U.S. startups (94,703 patents) over the 1980-2010 period.[3] Our main result is that startups with more disruptive potential are significantly more likely to go public and are less likely to exit by selling out. This result is robust after controlling for startups' age, size, financing rounds, market conditions, and other patent traits such as technological "breadth" (patents using vocabulary from diverse bodies of knowledge), similarity to other firms, patent citations, average word age, originality and fixed effects for startup cohorts, geographic locations, and technological categories.[4] Our results are also economically large as a one standard

[3]VC-backed startups are a relevant sample to study the role of technologies in exits. First, data is available to precisely link exit timing to patent data. Second, VC-backed startups account for a large fraction of the IPO and acquisition market (Ritter (2017)), produce a large share of total innovation (Gornall and Strebulaev (2015)), and understanding their exits is important given the importance of the VC market.

[4]Our results are also robust to estimations using logit, and multinomial logit; changes in the horizon over which we measure exits (ranging from one quarter to five years); and to focusing on the early part of our sample to limit potential truncation bias (Lerner and Seru (2017)). They are also present across all broad patent technological areas including biotechnologies and health science where Cunningham, Ederer, and Ma (2019) identify acquisitions motivated by preemption of future competition (i.e., "killer" acquisitions). They report that about 6% of acquisitions in their sample are likely "killer" acquisitions,

deviation increase in disruptive potential is associated with an increase of 25.2% in the probability of exit via IPO and a decrease of 18.8% in the incidence of sell-outs.

Conceptually, technologies with disruptive potential can either change the nature of *established* technology markets or they can lead to entirely *new* technologies. Although both are relevant and important, we divide patent vocabularies into established and new technological areas based on the age of each word (relative to when it first appears in the patent corpus). We then reexamine exits and find that disrupting established markets is far more important than new market creation in predicting IPOs and sell outs. These results strongly separate our study from the existing literature given that most existing studies focus on new market innovation.

Our second major finding is that aggregate economy-wide disruptive potential has declined substantially since the 1950s, and this trend accelerated in the 1990s. Although the overall trend is interrupted by occasional spikes during the 1970s (i.e., computers), the 1980s (i.e., genetics), and the 1990s (i.e., the internet), following each spike, the trend quickly reverts fully back to the long-term decline. As was the case for our main cross-sectional results for exits, this trend is also stronger for established technology spaces as we observe little change in the potential to disrupt by creating new technologies. These results are consistent with new ideas becoming harder to discover and develop (Jones (2009) and Bloom, Jones, Reenen, and Webb (2017)). Our findings indicate that this conclusion is particularly true in established markets.

The confluence of our first two conclusions (startups with disruptive potential exit via IPO, and disruptive potential has been declining) motivates a new technology-based explanation for the aggregate trends away from IPOs and toward sell-outs noted in recent studies. We assess this explanation by fitting a cross-sectional exit model (with and without our explanatory variables) over an initial period (1980-1995), and applying this model out-of-sample (1996-2010) to compute expected exit rates. A model that excludes our technological characteristics predicts an out-of-sample quarterly IPO rate of 0.84 percentage points. The actual rate was just 0.33 percentage points, confirming that IPOs

an economically important number but also one that indicates that our results are complementary and not mutually exclusive to those in their study.

were abnormally rare in recent years. Adding our new variables to the fitted model reduces this gap by roughly 20% overall, and by 40% for small IPOs (the market segment displaying the sharpest decline, see Gao, Ritter, and Zhu (2013)). A similar analysis reveals that we can explain an even larger 50% of the recent rise in sell-outs.

Our analysis adds to the recent literature offering explanations of the decline in IPOs and the rise in sell-outs.[5] Ewens and Farre-Mensa (2018) indicate that part of the IPO decline results from the increased bargaining power of founders whose preference is for control, and inexpensive private capital. Gao, Ritter, and Zhu (2013) suggest that lower IPO rates arise from changes in market structure that favor selling out (economies of scope). Doidge, Kahle, Karolyi, and Stulz (2018) argue that an increased focus on intangibles also plays a role. Our paper shows that changes in firms' technological traits can also account for part of the decline in IPOs and the surge in sell-outs.

Our findings also add to the literature studying the determinants of startup exits. The vast majority of past studies either examine IPO or sell-out exits in isolation or bundle them into a single proxy for successful exit (Bernstein, Giroud, and Townsend (2016) and Guzman and Stern (2015)). The small number that examine these exit choices jointly indicate that they depend on founders' private benefits of control, product market presence, and firms' growth potential (Cumming and Macintosh (2003), Bayar and Chemmanur (2011), Poulsen and Stegemoller (2008) or, Chemmanur, He, He, and Nandy (2018)). This limited evidence is surprising given the practical importance of exit payouts (Wang (2018) or Phillips and Zhdanov (2013)).

Finally, our paper complements recent studies using patent text to characterize technologies, especially contemporaneous work by Kelly, Papanikolaou, Seru, and Taddy (2019) who measure patent "importance" based on textual similarity to prior and future patents.[6] Our paper is unique both in terms of the economic question we address and also regarding measurement. Kelly, Papanikolaou, Seru, and Taddy (2019) study

[5] See Ritter and Welch (2002) and Lowry, Michaely, and Volkova (Forthcoming) for comprehensive surveys.

[6] Also see Packalen and Bhattacharya (2018) and Balsmeier, Assaf, Chesebro, Fierro, Johnson, Johnson, Li, Luck, O'Reagan, Yeh, Zang, and Fleming (2018), who identify new ideas using the first appearance of words and analyze their propagation over time.

long run technological change at aggregate and sector levels, whereas we focus on micro firm-level exit decisions and the trends in these exits over time. Methodologically, they measure importance as textual dissimilarity to previous patents and similarity to future patents. We focus on ex ante disruptive potential in established and new markets.

II Technological Characteristics and Exit Decisions

Our paper focuses on startup technologies and the decision to exit by either listing shares on the stock market or by selling out to other entities. We consider a simple rational framework in which a startup (i.e., the entrepreneurs and their backers) choose the exit method that maximizes the value received by equity holders upon exiting. This payoff is the total value conveyed to the startup by dispersed investors when startups exit through IPO and by strategic buyers when they exit by selling out. As a starting point, the theory of the firm (Grossman and Hart (1986) and Hart and Moore (1990)) suggests that this choice depends on whether the startup's value is higher as a stand-alone operating entity or when it is integrated with the assets of another entity. Our central hypothesis is that this dichotomy depends on startup technologies.

We hypothesize that the extent to which the startup's technology has either "disruptive" or "synergistic" potential is important. A technology has disruptive potential if it has the potential to eventually displace (i.e., reduce the value of) existing inventions and significantly influence the path of future innovation in its area (Dahlin and Behrens (2005)). In contrast, a technology has synergistic potential if its features complement (i.e., increase the value of) existing inventions and enhance the technological area. We propose that these features impact a startups' stand-alone value relative to its value as an acquisition target and hence they should predict exits.

By design, technologies with disruptive potential displace rather than improve existing technologies. Startups with such technologies can thus generate high valuations without relying on pairing these technologies with those of other firms, and hence their exit does not require any strategic assistance from other firms (Bayar and Chemmanur (2011)).[7]

[7] Relatedly, Darby and Zucker (2018) show that biotechnology firms go public when their innovations can be successfully commercialized, Chemmanur, He, He, and Nandy (2018) report that manufacturing

Maintaining independence can be optimal as it avoids the need to share rents with a potential acquirer, further increasing payoffs as initial equity holders can capture all rents. We thus predict that startups with disruptive potential should favor exits via IPO.[8]

In contrast, synergistic technologies have lower stand-alone values because their economic benefits arise primarily through combinations with existing technologies. In this case, exit via acquisition can achieve higher valuations (Higgins and Rodriguez (2006)) because it facilitates new innovation (Holmstrom and Roberts (1998)), technology coordination (Hart and Holmstrom (2010)) and synergies with existing technology portfolios (Cassiman and Veugelers (2006)).[9] In this situation, the gains to business combination can outweigh the costs of sharing future rents. We thus predict that startups with less disruptive potential and high synergistic potential should prefer exits via sell-outs.

If our cross-sectional exit predictions are stable over time, and the distribution of startup technological characteristics changes over time, our hypothesis also has predictions for the time series. If the disruptive potential is declining or synergistic potential is increasing, we would predict a structural decline in IPOs and an increase in sell-outs. Our hypothesis thus offers a potential new explanation for the aggregate decline in IPOs and the surge in sell-outs documented by recent studies (Gao, Ritter, and Zhu (2013)).

Existing research further motivates the importance of testing this hypothesis. Jones (2009) and Bloom, Jones, Reenen, and Webb (2017) document a secular decline in the productivity of technology research over time, and argue that breakthrough ideas are becoming increasingly difficult to find. This prediction is also consistent with product life cycle theories (Abernathy and Utterback (1978) and Klepper (1996)), which posit that markets mature and innovation becomes more incremental over time. Also providing motivation, Wang (2018) finds that entrepreneurs increasingly develop technologies that

firms favor IPOs when they have a defensible market share, and Poulsen and Stegemoller (2008) and Cumming and Macintosh (2003) show that higher growth potential tend to exit through IPOs.

[8] Relatedly, Hackbarth, Mathews, and Robinson (2014) show that the value of less developed growth options is higher in a stand-alone entity than in a combined entity, reinforcing our prediction if technologies with disruptive potential are less developed.

[9] Such technologies could also attract higher valuations in sell-out because buyers might have better information about synergistic technologies than disruptive ones (Kaplan (2000) or Hirshleifer, Hsu, and Li (2017)).

complement potential acquirers, which can increase sell-outs. Our second major hypothesis is thus that part of the recent shift in exits from IPOs to sell-outs might be attributable to the decline in disruptive potential we report.

III Data and Methods

A Patent Data and Text

We gather information from Google Patents for all 6,595,226 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 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 the Internet Appendix (Section IA.A).

[Insert Figure I about here]

The full text of each patent consists of three distinct sections: abstract, claims, and description. The claims section defines the scope of legal protection granted. The description section explicitly describes the characteristics of the invention/innovation. 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 (Packalen and Bhattacharya (2018) or Kelly, Papanikolaou, Seru, and Taddy (2019)).

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.[10] Following the literature, we eliminate common words appearing in more than 25% of all patents in a given year and rare words appearing only in one patent in a given year. Each patent j applied for in year t is represented by a vector $V_{j,t}$ of length N_t, and each element corresponds to the number of times the corresponding word is used by patent j. If patent j does not use a given word, the corresponding element is zero. This procedure ensures that all patents in a given year are represented by vectors in the same N_t-dimensional space.

Due to the large number of words used across all patents, these vectors are sparse. In 1980, the number of distinct words used in an average patent is 352 (median is 300), while there are N_t=400,097 distinct words across all patents. In 2000, the average and median are 453 and 338, and N_t is 1,358,694.

B Technological Disruptive Potential

We define the disruptive potential of a patent as its potential to change (potentially disrupt) the path of technological evolution in its area (Dahlin and Behrens (2005)). Our goal is to construct a variable that is theoretically motivated, measurable ex ante, and highly correlated with ex post evidence of disruption. Ex ante measurability is particularly important because our tests rely on predictive models of startup exits, which requires the exclusive use of ex-ante available data. Ex ante measures can also be important to practitioners and regulators who seek actionable information about startup valuation, job creation, and estimates of economic value likely created from innovation patterns.

To capture disruption empirically, we rely on the extent to which a given patent uses (new or established) vocabulary that is experiencing high growth in usage within the set of all contemporaneous patent applications. Intuitively, large surges in the use of existing vocabularies are likely associated with possible revolutions in the understanding of existing

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

ideas. Analogously, surges in new vocabularies likely indicates entirely new ideas with high potential. Because either is disruptive, our main specification measures disruptive potential based the intensity with which a patent broadly uses vocabulary that is surging in usage. However, we also consider separate measures based only on established or new vocabularies and we report later that our results are strongest for established technology areas. Importantly, all such measures can be computed solely using textual information available in the application year of each patent.

To implement this approach, we define an aggregate vector Z_t with elements containing the aggregate number of times a given word is used across all patent applications in year t. We then compute D_t, the normalized annual rate of change the vector Z_t, as[11]

$$D_t = \frac{Z_t - Z_{t-1}}{Z_t + Z_{t-1}}, \tag{1}$$

where division is element by element. The annual vectors D_t track the appearance, disappearance, and growth of specific technological vocabularies across all patents over time. Elements of D_t are positive if the corresponding word increases in use from year $t-1$ to t, and negative if it decreases.

[Insert Table I about here]

For example, Table I displays the ten words experiencing the largest increases and decreases in usage in specific years. In 1995, we detect increased use of terms related to genetics, such as "polypeptides", "clones", "recombinant" and "nucleic", following rapid progress in genome sequencing. In contrast, the terms "cassette," "ultrasonic," and "tape" are sharply decreasing. In 2005, the most rapidly growing words are related to the internet: "broadband", "click", "configurable", and "telecommunications".

To obtain the disruptive potential of a given patent j, we compute the frequency-weighted average of D_t over the words used by patent j:

$$\text{Disruptive Potential}_{j,t} = \frac{V_{j,t}}{V_{j,t} \cdot \overline{1}} \cdot D_t \times 100, \tag{2}$$

[11] We ensure Z_t and Z_{t-1} are in the same space by first mapping both to the expanded space associated with the union of the words either uses.

The operator "$\cdot$" denotes the scalar product and $\overline{1}$ denotes a unit N_t-vector. Intuitively, patents using words whose usage surges (indicated in D_t) have high disruptive potential. For example, Table I indicates that patents using "polypeptides", "clones", "recombinant" and "nucleic" in 1995 (but not necessarily in other years) would be classified as having high disruptive potential. Analogously, a patent using obsolete vocabulary such as "cassette," or "tape" in 1995 would have low (possibly negative) disruptive potential. Despite this intuition, not all inventions with disruptive potential succeed (Christensen (1997)). Our goal is to identify a well-microfounded predictor that performs well. Section IV (validation) shows that our measure strongly predicts ex post disruptive outcomes for a patent, including citations, structural breaks in citation patents, economic value, appearance on lists of influential patents, and the incidence of established firms disclosing concerns about disruption.

C Technological Breadth and Similarities

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

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

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

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

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

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. Specifically, we form the aggregate private firm (foreign firm) word vectors and compute the cosine similarity between each patent j and these aggregate vectors.[12]

D Linking Patent-Level Traits 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 attention to startups that are granted at least one patent during the sample period. To link patents to startups, we follow Bernstein, Giroud, and Townsend (2016) and develop a fuzzy name matching algorithm (see Section IA.B in the Internet Appendix).[13] A startup enters our startup-quarter sample in the quarter it is founded (based on founding dates in VentureXpert)

[12]Because these groups contain very large numbers of patents, we aggregate them over just one year t.

[13]Lerner and Seru (2017) note that bias can occur in matching patent assignments to startups because patents can be assigned to subsidiaries with different names than their parent corporations. However, this issue is limited in our sample as startups are small and are unlikely to have complex corporate structures.

and exits the sample when its final exit is observed on the "resolve date" from VenturеXpert. Startups still active in November 2017 are unresolved.[14] We exclude startups if their founding date is missing or is later than the resolve date. Our final sample contains 347,929 startup-quarter observations, corresponding to 9,167 unique startups and 94,703 patent applications.

We aggregate patent-level technological characteristics (text-based and others) for each startup-quarter using depreciated sums over the past 20 quarters and a quarterly depreciation rate of 5%. For example, the technological disruptive potential of startup i in quarter q corresponds to the depreciated sum of the disruptive potential of all its patent applications in the past five years, normalized by the number of patents startup i applied for over that period.[15] We define the exit variables (IPO or sell-out) as binary variables equal to one if startup i experiences a given exit in quarter q. Variable constructions are explained in detail in Table A1.

E Descriptive Statistics

Table II presents descriptive statistics for our new text-based technological characteristics as well as existing patent variables from the literature. All variables are defined in Table A1. We present patent-level statistics for the full sample of patents (1930-2010) in Panel A, and startup-quarter-level statistics (1980-2010) in Panel B. Focusing on our central new variable – technological disruptive potential – we note that its empirical empirical distribution is highly skewed. The first row of Panel A indicates that the average disruptive potential of patents is 1.64, 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. Despite the aggregation of their patents, we observe a similar asymmetry in the disruptive potential of startups, with a median of zero, and the 75^{th} percentile is 0.98.

[14] Ewens and Farre-Mensa (2018) note that unresolved firms can result from stale data collection, and we code firms as failed if seven years pass since their last funding round.

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

[Insert Table II about here]

Table II also provides statistics for our other text-based measures of patent characteristics. Unlike technological disruptive potential, patent breadth is more evenly distributed, indicating less skewness in technological specializations. We also observe some variation in similarity across patents, but the overall levels are low, which is not surprising given the large range and diversity in the vocabulary used across all patents. Overall, the patent and startup-quarter statistics are similar, indicating that the technological characteristics of VC-backed startups are roughly representative of those in the economy at large. Relevant for our regression analysis, Panel B further indicates that the quarterly IPO rate (i.e. the number of IPOs in a quarter divided by the number of active startups in that quarter) is 0.42 percentage points, and the quarterly sell-out rate is 0.73 percentage points.[16]

IV Validation of Disruptive Potential

We consider three tests of validation based on (1) patent-level ex post citation patterns and economic value, (2) appearance on a list of unambiguous breakthrough patents from USPTO, and (3) direct mentions of disruption concerns by publicly traded firms operating in markets related to our startups.

A Citation Patterns and Economic Value

As disruptive patents should become highly cited, we first consider the (logarithm of one plus the) number of citations each patent receives ex post on Google Patents.[17] Second, and more important, we expect disruptive patents to trigger structural breaks in the citation patterns of related patents going forward. We use the measure of structural breaks developed by Funk and Owen-Smith (2016) (mCD), which is based on whether the patents citing a focal patent also cite the patents cited by the focal patent (see

[16] We report additional information about the sample firms in the Internet Appendix in Table IA1. Relative to the founding date, IPOs and acquisitions play out over time. Of these, IPOs occur fastest on average. The average firm applies for its first patent after 4.42 years and receives its first round of VC funding 5.29 years after its founding. All of these numbers are mechanistically reduced when measured relative to the first patent instead of the founding year.

[17] We focus on citations received within five years of the grant date to limit truncation bias, but results are robust to using citations through 2013 or just those in the first year or two after grant.

also Wu, Wang, and Evans (2019)). Third, following Kogan, Papanikolaou, Seru, and Stoffman (2016), we examine the economic value of patents as disruptive patents should be valuable. This final test is restricted to patents assigned to publicly listed firms as it requires observing stock market reactions to each patent's issuance.

[Insert Table III about here]

Table III presents the results of these three patent-level tests where each ex post metric is regressed on each patent's ex ante disruptive potential. We include cohort fixed effects to isolate variation across patents granted in the same year, and we cluster standard errors by cohort. Panel A shows that our measure of disruptive potential is positively associated with all three ex post metrics. These findings are highly statistically significant with t-statistics ranging from 4.35 to 7.48. Patents with more ex ante disruptive potential create more impact in the form of citations, trigger more structural breaks in technological evolution, and generate more economic value in the stock market.

In Panel B, we additionally control for heterogeneity in the citation patterns and unobservables across technology areas as we include interactions between cohort and technology area fixed effects based on two-digit NBER technology codes (Hall, Jaffe, and Trajtenberg (2001)). In Panel C, we additionally include four patent characteristics as controls including technological breath and similarities between each patent and the set of patents granted to private, public, and foreign firms. We continue to observe strong and positive relationships between these ex post outcomes and our measure, consistent with our measure capturing disruptive potential.[18]

B Unambiguous Breakthrough Inventions

We consider a collection of twelve breakthrough patents, as identified by the USPTO's "Significant Historical Patents of the United States".[19] Panel A of Table IV displays these patents and we score each based on its disruptive potential percentile rank in its grant year.

[18]The associations in Table III also indicate that ex post outcomes are not perfectly predictable, something we expect of any ex ante predictive measure. This finding echoes Christensen (1997) and Christensen and Rosenbloom (1995)).

[19]Breakthrough Historical patents before 1960 are from http://www.uspat.com/historical/. Recent breakthrough patents are noted for the revenue they generated.

For instance, a value of 0.95 indicates that the patent is in the top 5% of the distribution. Reassuringly, we find high disruptive potential for these breakthrough patents at the time of their application to the USPTO, as they collectively rank at the 82nd percentile of disruptive potential. The patents that displayed the highest disruptive potential are the "Complex computer" in 1944 (#2668661) and DNA modifications in 1980 (#4399216). Other key inventions, such as the satellite (#2835548), laser (#2929922), and PageRank (#6285999), also use vocabulary with high disruptive potential.[20]

[Insert Table IV about here]

Second, we consider a list of 101 important patents between 1930 and 2010 identified by Kelly, Papanikolaou, Seru, and Taddy (2019) based on several on-line lists, and we report disruption percentile scores for the full list in the Appendix (Table A2). We present more parsimonious summary statistics in Panel B of Table IV and find that the disruptive potential of these patents is in the 71st percentile on average, and the median is in the 81st percentile.[21] These average percentiles are measured rather precisely as the standard error is 0.027. Overall, these findings confirm that the vast majority of breakthrough patents have high levels of our ex ante measure of disruptive potential.

Importantly, Table A2 also confirms that our measure is distinct from the importance measure developed by Kelly, Papanikolaou, Seru, and Taddy (2019). For instance, our measure classifies patents associated with color film (#2059884), the process of fluid catalytic cracking (#2451804), and the cellular telephone (#3906166) as highly disruptive. However, they do not score highly using the Kelly, Papanikolaou, Seru, and Taddy (2019) measure since they are not highly dissimilar to previous patents. Yet, these inventions

[20] 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 based on citations. Yet, our new measure classifies these patents as highly disruptive. Kelly, Papanikolaou, Seru, and Taddy (2019) similarly note that some patents classified as significant based on their measure attract few citations, and provide illustrative examples. One example is patent #174465 issued to Graham Bell for the telephone in 1876 having received only 10 citations until March 2018.

[21] For comparison, Kelly, Papanikolaou, Seru, and Taddy (2019) report that the average patent is in the 84th percentile of the distribution using their ex-post patent significance measure, and the KPSS measure of patent value for these patents is in the 68th percentile. Thus, despite being computed using ex ante information, our measure of disruptive potential overlaps some with that of Kelly, Papanikolaou, Seru, and Taddy (2019), which utilizes ex post information (i.e., textual similarity with future patents).

undeniably disrupted *established* technological areas.

C Perceived Disruption Risk

As a third test of validation, we consider publicly traded firms operating in related product markets to each startup, as measured by identifying the 25 publicly traded companies having 10-K business descriptions (Hoberg and Phillips (2016)) with the highest cosine similarity to the product description text of each startup (from Venture Expert based on the date of first funding) in each year. The sample for this test is limited to 1996 to 2010 due to electronic 10-K data availability, and includes 5,417 distinct startups (60% of our original startup sample).

For each publicly traded peer for each startup, we compute the fraction of paragraphs in its 10-K that mentions words related to technology-based disruption using three textual queries. First, we summarize mentions of technological change as paragraphs having word roots "technol" and "change". Second, we summarize mentions of technology competition as paragraphs having the word roots "technol" and "compet". Third, we summarize paragraphs having the more strict set of paragraphs containing the roots of "technol" and "compet" together with either "disrupt", "change", or "obsoles". Finally, we average each of these three measures across the 25 public peers for each startup.

[Insert Table V about here]

Finally, we regress these three measures on the focal startup's disruptive potential measured at the time of the startup's first funding round. Table V displays the results and shows a positive link between disruptive potential and all three public-peer disruption mentions. These specifications include either year fixed effects or a more complete set of fixed effects including year, technology, startup location, startup age, and startup cohort as noted in the table. These results further support the validity of our measure of ex ante disruptive potential and also its economic relevance given the importance of public firms.

V Disruptive Potential and Startup Exits

A Main Results

In this section, we examine the link between disruptive potential and startup exits. In particular, we estimate cross-sectional regressions based on our startup-quarter panel database (see Section III.D). We consider dependent variables including the IPO exit and sell-out exit dummy in quarter q, whereas our explanatory variables are ex-ante measurable in quarter $q-1$. Our main specification reports estimates using the competing risks hazard model of Fine and Gray (1999) and the linear probability model with fixed effects. However, our results are robust to alternatives including the logit, multinomial logit, and Cox hazard models.[22] We cluster standard errors by startup.

[Insert Table VI about here]

Table VI displays the results and the first two columns show the competing risk model. Column (1) indicates a strong positive link between a startups' disruptive potential and the likelihood of exiting via IPO in the next quarter. The point estimate is 0.252 with a t-statistic of 13.09. In contrast, column (2) shows that sell-out exits are negatively related to disruptive potential with a cofficient of -0.188 and a t-statistic of -7.55.[23] These results are not only statistically significant, the coefficients also indicate economically large impact: a one standard deviation increase in disruptive potential is associated with a *proportional* 25.2% increase in the quarterly IPOs exit rate and an 18.8% decrease in the sell-out rate.

Columns (3) and (4) of Table VI display results for the linear probability models. Although this approach ignores potential dependence across exits (i.e., competing risks), it allows us to include a wide array of fixed effects including: founding year cohort, year of

[22]Similar to a Cox hazard model, a competing risks hazard model explicitly models the "risk" of choosing a particular exit. However, unlike a Cox hazard model, it accounts for multiple potential hazards that are mutually exclusive. See Avdjiev, Bogdanova, Bolton, Jiang, and Kartasheva (2017) for a recent study using competing risk models in finance.

[23]In Internet Appendix Table IA2, we show that these results continue to hold if we only use acquisition exits that are clearly successful (larger than $25 million in 2009 dollars) to avoid the potential misclassification as highlighted by Maats, Metrick, Yasuda, Hinkes, and Vershovski (2011). We thank Josh Lerner for suggesting this test.

first VC funding cohort, state, operating year and technological area fixed effects.[24] The table shows that our statistical inferences and our economic magnitudes are very similar to those for the competing risks model in the first two columns.[25]

Table VI also shows that other text-based technological characteristics predict startup exits. For example, startups' technological breadth is positively related to IPO incidence and negatively related to sell-out incidence. A likely explanation is that high breadth technologies are less redeployable and have lower synergies (e.g., Bena and Li (2014)). Hence, IPO is a more sensible exit. We also find that startups whose patents are more similar to those of other private firms are significantly less likely to exit through sell-outs (t-statistic of -8.62) and are marginally more likely to go public (t-statistic of 1.70). These results are in line with the negative link between product market similarity and acquisitions for public firms documented in Hoberg and Phillips (2010). Finally, firms holding patents that are more similar to that of lead innovators are significantly more likely to go public (t-statistic of 4.58).

Following the literature, we also include controls for market activity based on relative valuation (*Log(MTB)*) and stock returns (*MKT Return*), and an identifier for the last quarter of the year (Lowry (2003) and Pastor and Veronesi (2005)). Consistent with earlier research, we find that startups are more likely to exit via IPO after periods of strong stock market performance. We also include startup technology controls including the log of one plus the number of patent applications over the past five years, a dummy variable indicating zero applications in the past five years, the originality of startup's patents, and patent citations.[26] The zero patent dummy indicates that startups with a positive number of patents are more likely to sell-out than go public. The total patents variable offsets this effect, as startups with higher overall patent counts are associated

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

[25]Because the overall IPO rate in the sample is 0.42%, the 25.2% economic impact from column (1) indicates an increase of 0.106 in the probability of an IPO. This is close to the 0.082 estimate in column (3).

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

with a shift towards IPOs.[27] Finally, more future citations (originality) are positively (negatively) associated with exits via both IPOs and sell-outs.

B Established and New Technology Spaces

Conceptually, technologies with disruptive potential may significantly alter the nature of innovation in either *established* technological areas (Tushman and Anderson (1986)), or entirely *new* areas (Acemoglu, Akcigit, and Celik (2014)). To assess which dimension matters, we consider refined measures of disruptive potential based on either established vocabulary or vocabulary newly appearing in the patent corpus.

Our measure for established technological markets focuses on words that have been used in the patent corpus for at least ten years. We interpret these older words as pertaining to established technology spaces and define disruptive potential for these technology areas as:

$$\text{Disruptive Potential (Established)}_{j,t} = \frac{V_{j,t}^{10+}}{V_{j,t}^{10+} \cdot 1} \cdot D_t^{10+} \times 100, \tag{5}$$

where V^{10+} and D^{10+} are the vectors V and D defined in Section III.B, except we remove elements relating to words less than ten years old within the patent corpus at time t. By construction, this measure is not influenced by young or new words. For example, patents containing the word "internet" in 1993 will not necessarily score highly on this measure because the new term "internet" is not part of the established vocabulary. Instead, patents will score higher when they use older words whose usage suddenly surges in volume across all patent applications in a given year. Such patents thus belong to "second (or later) waves" of innovation within a specific established technology space.[28]

Analogously, to measure the potential of a patent to create entirely new technology areas, we consider only words first observed in the patent corpus during the most recent

[27] For example, the estimates in column (3) indicate that, compared to a startup with no patents, a startup with one patent is 0.35 p.p. less likely to go public while those with seven or more patents are more likely to go public.

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

ten years. We define disruptive potential for these new technology areas as:

$$\text{Disruptive Potential (New)}_{j,t} = \frac{V_{j,t}^{<10}}{V_{j,t}^{<10} \cdot \mathbf{1}} \cdot D_t^{<10} \times 100, \tag{6}$$

where $V^{<10}$ and $D^{<10}$ are the vectors V and D as before, but we only keep the elements relating to words less than ten years old at time t. Intuitively, the above two measures form a complete decomposition of our main variable of disruptive potential into two components: potential to disrupt existing and new technology areas.[29]

[Insert Table VII about here]

To examine the relationship between exits and these two disruptiveness components, we aggregate these two patent-level measures to firm-quarter variables (analogously to our original variable) and we then include both components in our baseline regressions. Table VII presents the results for the competing risk and OLS linear probability models. We find that IPO exits are strongly related to startups' potential to disrupt established areas (with t-statistics of 10.20 and 5.54), but not to their potential to create new areas (with t-statistics of 0.99 and -0.90). Similarly, the incidence of sell-outs is more strongly negative for established markets although both components are significant. We conclude that disruption in existing technological areas is more important in determining startup exits. This result further suggests that the ability to conduct business as a stand-alone entity (and exit via IPO) is particularly valuable when disruption gains potentially come at the expense of existing market participants.

C Post-IPO Predictions

We consider the subsample of our startups that exit via IPO and test auxiliary predictions. Our first prediction is that more disruptive startups need to raise more capital to expand quickly, and hence a large fraction of their IPO shares should be primary shares (i.e., fewer

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

secondary shares). The first two columns of Table VIII test this hypothesis. We find that more disruptive startups indeed have a *lower* fraction of secondary shares (column (1)) and are less likely to include any secondary shares at all (column (2)) in their offering.

[Insert Table VIII about here]

Our remaining tests are based on three text-based product characteristics from 10-K filings, which only becomes available for startups that go public after 1997: industry concentration (HHI), total product similarity of publicly traded peers (TSimm), and product market fluidity (see Hoberg and Phillips (2016) and Hoberg, Phillips, and Prabhala (2014)). The life cycle model of Abernathy and Utterback (1978) suggests that our most disruptive startups are likely in the earliest stage of the life cycle (product development), a stage where high competition and rapid product turnover (high product market fluidity) are expected. Table VIII confirms these predictions as newly-public firms with more technological disruptive potential indeed exit into more contested and fluid markets. High fluidity, in particular, indicates a high susceptibility of these markets to disruption as this is direct evidence that products in these markets are more easily revised. In contrast, IPO firms with less disruptive potential exit into more stable markets with less competition, and higher levels of product differentiation.

D Robustness

In Table IX, we use the same specifications as reported in Table VI, but we further control for the startup's financing, as previous research reports that the amount of VC funding (a proxy for startups' implied valuation and outsiders' ownership) predicts startups' exits (Cumming and Macintosh (2003)). These tests are important because the link between disruptive potential and startup exits could reflect the different funding needs of startups developing disruptive innovation, or the different incentives for IPOs (i.e., larger payoffs) for VC funds allocating more capital to startups displaying more disruptive potential.[30] To account for the possible role of funding, we include two additional controls: a startup's cumulative prior VC funding (from founding to quarter $q-1$) and a dummy variable

[30]We confirm this funding intuition in Table IA3 of the Internet Appendix.

indicating whether a startup received funding in the last five years. Table IX confirms that the financing variables are strong predictors of startups' exits. However, our inferences regarding technological disruptive potential are fully robust, indicating that our findings likely cannot be explained by financing.

[Insert Table IX about here]

Table X shows that the results are robust to examining longer exit horizons. We increase the measurement window for exits from one quarter to five years using increments of one year and focus on our baseline OLS specification that includes all fixed effects as described above. We only report coefficients for our text-based technology variables for conserve space. Panel A shows that the positive relationship between disruptive potential and IPO exits remains strong at all horizons and Panel B shows that the negative relation between disruptive potential and sell-outs persists for one year and fades after two.

[Insert Table X about here]

Panel C of Table X examines whether technological traits are related to the propensity to remain private for longer periods of time (horizons from one quarter to five years). This analysis is motivated by Gao, Ritter, and Zhu (2013) and Ewens and Farre-Mensa (2018), who suggest that startups are remaining private longer. The results show that startups with higher disruptive potential exit more quickly on average and hence remain private for shorter periods relative to less disruptive startups.

We report additional robustness tests in the Internet Appendix. Table IA2 shows that our results hold when standard errors are clustered by year, startup technology, or startup cohort. The results also hold when we exclude citations as a control, when we only control for the size of a startup's patent portfolio, and in a subsample comprised only of startup-quarters that resulted in exits (where the dependent variable equals one if the exit is an IPO and zero if it is a sell-out). This latter result confirms that our results are not mechanically driven by IPO firms potentially staying private longer, as each firm only has one observation in this test. Finally, Table IA4 shows that our results are stable across sub-periods. This ensures that our results are not driven by either (A) startups with "unresolved" status or (B) truncation bias associated with patents not yet granted (Lerner and Seru (2017)).

VI The Evolution of Disruptive Potential

Our results thus far illustrate a strong cross-sectional relationship between startups' exit decisions and their technological disruptive potential. This section studies the aggregate evolution of patents' and startups' disruptive potential.

A Patents' Disruptive Potential in the Last Century

We compute aggregate disruptive potential using an average based on a 20-quarter rolling window.[31] Panel A of Figure II plots the result from 1930 to 2010 annualized using a four-quarters moving average. The figure shows transitory periods of sharply increasing disruptive potential with an initial peak around 1950 at a level that is roughly double that in 1930. The period around 1950 is often viewed as 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]

Despite these periodic surges in disruption, the 1930-2010 period is characterized by a protracted and steady long-term decline in the disruptive potential of U.S. patents. Between 1950 and 2010, the average disruptive potential of patents 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 technol-

[31] We first sum disruptive potential over the patents applied for in a given quarter. We then apply a 5% quarterly rate of depreciation over the 20 rolling quarters and then divide by the number of patents applied for in these 20 quarters to arrive at the average disruptive potential over time.

ogy and location fixed effects (unreported for brevity). Rather, the decline indicates a widespread deceleration in vocabulary usage growth rates among U.S. patents.[32]

Panel B of Figure II separately plots the evolution of the two components of patents' disruptive potential, defined in Section V.B. We note that most of the variation in patents' disruptive potential comes from the potential to disrupt established technological areas. Indeed, the aggregate potential of patents to create new areas has remained stable until the eighties, increasing slightly in the 1990s, and declining after 2000. Taken together, the secular decline in patents' disruptive potential echoes recent research highlighting the increasing difficulty to generate new ideas (e.g., Jones (2009) and Bloom, Jones, Reenen, and Webb (2017)). Our findings suggest that the increased difficulty of finding new ideas is particularly salient in established markets.

B Disruptive Potential, IPO, and Acquisitions

We next contrast the decline in patents' disruptive potential to the aggregate evolution 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). To facilitate comparison, Panel C displays the evolution of aggregate patents' disruptive potential during the same

[32] To conserve space, we display and discuss our other text-based variables (the aggregate evolution of patents' technological breath and similarity to groups of patents) in the Internet Appendix.

period. The evolution of IPO activity rather closely maps the aggregate dynamics of disruptive potential during this thirty-year period. The number of IPOs drops around 1990, coinciding with a decline in disruptive potential that follows the earlier surge in genetic science in the mid-1980s. There were more IPOs as the nineties progressed, when disruptive potential experienced a very large increase. The decline in IPO intensity then began in the early 2000s, when the average disruptive potential of U.S. patents also started to plummet. Although much of the variation appears linked to disruptive potential, the rise in exits around 2005 appears unrelated, and might be linked to increased private equity activity during this time.

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 technology bubble and the financial crisis. Yet, the number of acquisitions remains significantly higher since the mid-nineties when compared to the 1980-1995 period, suggesting a relationship between the surge in aggregate acquisitions and the decline in disruptive potential of U.S. patents. Although the aggregate pattern for sell-outs is less striking than that for IPOs, it suggests that acquisitions tend to be high when overall disruptive potential is lower.

[Insert Figure IV about here]

In Figure IV, we display the evolution of disruptive potential along with IPO and sell-out rates for the startups in our sample. For the sake of comparison, we compute the aggregate stock of each variable for the set of patents granted to startups as in Figure II. In addition, we scale the aggregate quarterly number of IPOs and acquisitions by lagged real GDP to obtain aggregate exit rates. Reassuringly, the trends in our startup sample closely map those of the aggregate dynamics, indicating that the technological changes and exit patterns occurring among startups is mirroring economy-wide changes. We again observe that periods with elevated disruptive potential exhibit more intense IPO rates, and lower sell-out rates.[33]

VII Disappearing IPOs and Surging Sell-Outs

We now estimate whether (and how much of) the recent shift in exits away from IPOs toward sell-outs can be attributed to changes in technological characteristics.

A Prediction Errors for Startups' Exit

To examine the impact of changing technological characteristics on startups' exit choice, we use methods from the disappearing dividends literature (see Fama and French (2001) and Hoberg and Prabhala (2009)) and proceed in two steps. First, we estimate two linear probability models (a "Base" model and a "Text" model) using quarter-by-quarter Fama and MacBeth (1973) regressions where the dependent variable is the incidence of IPO exits in each quarter during the initial period 1980-1995 (the "pre-period"). The base model's independent variables are from the existing literature and include the (log) of the startups' age and the startup's patent stock. The Text Model adds to this startups' disruptive potential and the additional text-based technology traits considered in this study. Second, we compute predicted values of IPO incidence for each startup-quarter in the 1996-2010 period (the "post-period") by using the average coefficients estimated in the pre-period as a predictive model, and applying this predictive model using the actual values of the independent variables in the post-period. We then average the predicted IPO rates across all startups in each quarter and compare them to the actual observed quarterly IPO rates in the post-period. Since the coefficients are locked in at their pre-period fitted values, we are able to isolate variation in these predicted IPO rates in the post-period that is due to changing startups' characteristics. We repeat these steps for sell-outs to compare actual and predicted sell-out rates.

[Insert Table XI about here]

[33] Notably, the overall rate of exit (i.e., IPO and sell-out) has remained stable between 1980 and 2010 at about 0.9%, except between 2002 and 2006 where it dropped to 0.7%. This stability indicates that our analysis is not biased by changes in overall exit trends.

Test 1 in Panel A of Table XI indicates that the base model yields an average predicted quarterly IPO rate of 0.84 percentage points in the post-period. This predicted incidence is substantially higher than the actual IPO rate, which is 0.33 percentage points per quarter in the post-period, implying a prediction error of 0.51. The predicted IPO rate is thus 2.5 times higher than the actual rate, confirming that observed IPO rates in the post-period are "abnormally" low. Using the Text model, the average predicted IPO quarterly rate in the post-period declines to 0.75 percentage points, which is still higher than the actual incidence rate as the prediction error is 0.42. In the rest of Table Table XI we modify the definition of the pre- and post-periods or to the forecasting horizon considered (i.e., the lags between the dependent and independent variables). Although a significant portion remains unexplained, our overall finding across the array of specifications shown in the Panel A is that changes in technological characteristics account for roughly 20% of the recent decline in IPO rates.

Panel B reports parallel analysis for sell-out rates. A benchmark linear model that excludes our technology variables estimated in the pre-period yields an average predicted sell-out incidence of 0.60 percentage points per quarter in the post-period. Compared to the actual rate of 0.86 per quarter, the base model's prediction is 42% lower than the actual rate, suggesting that the prevalence of sell-outs in recent years is "abnormally" high. Using the Text model, the prediction gap narrows significantly, as we obtain a predicted sell-out rate of 0.75 percentage points per quarter. When we alter specifications in the remainder of Panel B, we observe that changes in startups' technological characteristics explain between 26% and 71% of the surge in sell-outs. We conclude that roughly 50% (the average across specifications) of the surge in trade sales is accounted for by changes in startups' technological characteristics.

B Small and Large IPOs

The recent dearth of IPOs is particularly pronounced for smaller-company IPOs (see Gao, Ritter, and Zhu (2013) and Doidge, Kahle, Karolyi, and Stulz (2018)). To further validate the role of technological changes in explaining the recent decline in IPOs, we rerun the analysis in Table XI separately for small and large IPO exits. We measure IPO size using pre-IPO sales data from Gao, Ritter, and Zhu (2013) and inflation adjusted to 2009 dollars. We define an IPO as "small" if its pre-IPO sales are below the median in our sample ($25 million), and as "large" if its pre-IPO sales exceeds that amount. We then examine each subsample and estimate the probability that a given startup exits through a small (large) IPO in a given quarter in the pre-period quarter-by-quarter. As before, we estimate the model with and without our text-based technological variables, and compare the predicted IPO rate in the out-of-sample period to the actual rate.

[Insert Table XII about here]

Table XII displays the results. Panel A indicates that, across six specifications which vary the definition of the pre- and post-periods or the forecasting horizon, changes in startups' technological traits account for roughly 37% of the disappearing small IPO anomaly in the recent period. In sharp contrast, Panel B reveals that adding startups' technological characteristics to the regression models (estimated in the pre-period) does not bring the average predicted rate of large IPOs closer to its actual value in the post-period. We conclude that changes in startups' technological characteristics are particularity important to explaining the decline in small IPOs.

C The Role of Product Market Stability

We also explore the role of product market maturity in explaining the recent shift from IPOs to sell-outs. We posit that markets reaching maturity (e.g., markets that effectively reached a dominant product design) are likely to experience the most extreme decline in IPO rates. In these markets, breakthrough inventions obtain only with very high search costs (Jones (2009) and Bloom, Jones, Reenen, and Webb (2017)) as the best ideas are already "picked over", so that the decline of startups' disruptive potential should be a stronger predictor of lower IPO activity. To test this idea, we follow Hoberg, Phillips, and Prabhala (2014) and compute the degree of product market fluidity in each startup's product market from 1980 to 2010 using the business description text that is available at the time of the first funding round in Thomson Reuters's VentureXpert. We first compute the aggregate change in product description vocabulary used by startups as the year-over-year change in the frequency of word 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 this 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 startups receiving their first money between 1980 and 2010.

[Insert Table XIII about here]

To assess whether changes in startups' technological traits account for the recent decline in IPOs differently in stable and unstable markets, we divide our startup-quarter observations into above and below median fluidity subsamples, based on median breakpoints chosen separately for each cohort of startups (based on the year of the first funding round) and repeat the prediction procedure discussed above across each subsample. Panel A confirms that startups operating in stable markets are less likely to exit via IPO relative to startups in fluid markets, with IPO rates of 0.30% and 0.37% per quarter in the post-period (1996-2010). Moreover, changes in startups' technological characteristics explain roughly 25% of the dearth of IPOs in stable markets. This figure is tightly estimated across different specifications. In contrast, changes in startups' technological attributes account for just 3% of the dearth of IPOs in fluid markets. This figure ranges between -7% and 15% across different specifications. Although market stability is relevant for understanding the evolution of IPOs, Panel B indicates that such stability has little effect in moderating our ability to explain the surging sell-outs anomaly, as the average improvement is 50% and 49% in stable and fluid markets, respectively.

VIII Conclusions

We develop new ex ante measures of technological disruptive potential and other technology characteristics using textual analysis of 6,595,226 U.S. patents from 1930 and 2010.

We find that startups with higher ex-ante disruptive potential are more likely to exit via IPO, and are less likely to exit via sell outs. One particularly novel aspect of our conclusions, which is not examined in existing studies, is that our results are strongest when we specifically focus on the ability to disrupt more established technology markets rather than the ability to disrupt by creating entirely new technologies.

The economics of disruptive potential are most intuitive when juxtaposed against the alternative of synergistic potential. This juxtaposition intuitively is one of substitutes versus complements. Startups with disruptive potential likely favor IPOs because disruptive technologies tend to be economic substitutes for existing technologies and thus lack synergies for buyers. Additionally, these technologies likely enable startups to establish independent markets, allowing their owners to extract rents without having to share with a potential acquirer. In contrast, patents with synergistic potential tend to have high complementary value when combined with existing technologies. Startups owning these patents likely favor sell outs because the synergies facilitate increased value creation that overcome the need to share the gains with an acquirer.

Our second major finding is that technological traits have changed dramatically over time and we document an economy-wide decline in technological disruptiveness that began after World War II. Because our central thesis is that startups with disruptive potential are more likely to exit via IPOs, it follows that the aggregate decline in disruptive potential we document might also explain the recent aggregate decline in IPOs and the surge in sell-outs. We estimate that roughly 20% of the decline in IPOs can be attributable to changes in technological traits. Analogously, roughly 50% of the surge in sell-outs can be explained by these same traits.

References

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

Acemoglu, Daron, Ufuk Akcigit, and Murat Alp Celik, 2014, Young, restless and creative: Openness to disruption and creative innovations, Working Paper 19894 National Bureau of Economic Research.

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.

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

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

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

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.

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

Chemmanur, Thomas, 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.

Christensen, Clayton, 1997, *The innovator's dilemma: When new technologies cause great firms to fail* (Harvard Business School Press: Boston).

———, and Richard Rosenbloom, 1995, Explaining the attacker's advantage: Technological paradigms, organizational dynamics, and the value network, *Research Policy* pp. 233–257.

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

Cunningham, Colleen, Florian Ederer, and Song Ma, 2019, Killer acquisitions, Working Paper.

Dahlin, Kristina, and Dean Behrens, 2005, When is an invention really radical? Defining and measuring technological radicalness, *Research POlicy* pp. 717–737.

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?, *Journal of Applied Corporate Finance* pp. 8–16.

Doidge, Craig, Andrew Karolyi, and Rene Stulz, 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, Working Paper.

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

Fama, Eugene, and J. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 71, 607–636.

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.

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

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.

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

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

Hackbarth, Dirk, Richmond Mathews, and David Robinson, 2014, Capital structure, product market dynamics, and the boundaries of the firm, *Management Science* 60, 2971–2993.

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

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

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

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

Higgins, Matthew, and Daniel Rodriguez, 2006, *Journal of Financial Economics*pp. 351–383.

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

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

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

——— , 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.

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

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

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

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

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

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

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

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

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

———, Roni Michaely, and Ekaterina Volkova, Forthcoming, Initial public offerings: A synthesis of the literature and directions for future research, *Foundations and Trends in Finance*.

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

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

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

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

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

Ritter, Jay, 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.

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

Wang, Xinxin, 2018, Catering innovation: Entrepreneurship and the acquisition market, Working Paper.

Wu, Lingfei, Dashun Wang, and James Evans, 2019, Large teams develop and small teams disrupt science and technology, *Nature* pp. 378–382.

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)

FIG. 1 FIG. 2 FIG. 3

Classifications

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

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

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

Y10S707/99937 Sorting

Hide more classifications

US6285999B1

US Grant

Download PDF Find Prior Art

Inventor: Lawrence Page

Current Assignee: Leland Stanford Junior University , Google LLC

Original Assignee: Leland Stanford Junior University

Priority date: 1997-01-10

Family: US (10)

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

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

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

Description

CROSS-REFERENCES TO RELATED APPLICATIONS

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

STATEMENT REGARDING GOVERNMENT SUPPORT

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

FIELD OF THE INVENTION

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

BACKGROUND OF THE INVENTION

Due to the developments in computer technology and its increase in

Claims (29)

What is claimed is:

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

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

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

processing the linked documents according to their scores.

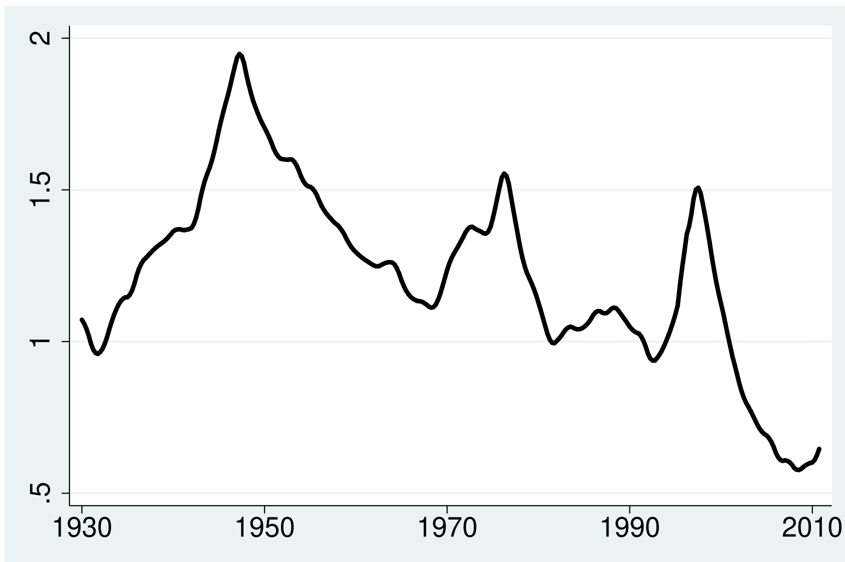
2. The method of claim **1**, wherein the assigning includes:

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

Figure II: Time Series of Aggregate Disruptive Potential

This figure reports the evolution of *Disruptive Potential* for the aggregate patent corpus from 1930 to 2010. *Disruptive Potential* is defined at the patent level in Section III.B and Equation 2. *Disruptive Potential (Established)* and *Disruptive Potential (New)* are defined at the patent level in Equations 5 and 6, respectively. 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. All series are reported as four quarter moving averages.

Panel A: Disruptive Potential



Panel B: Decomposition of Disruptive Potential

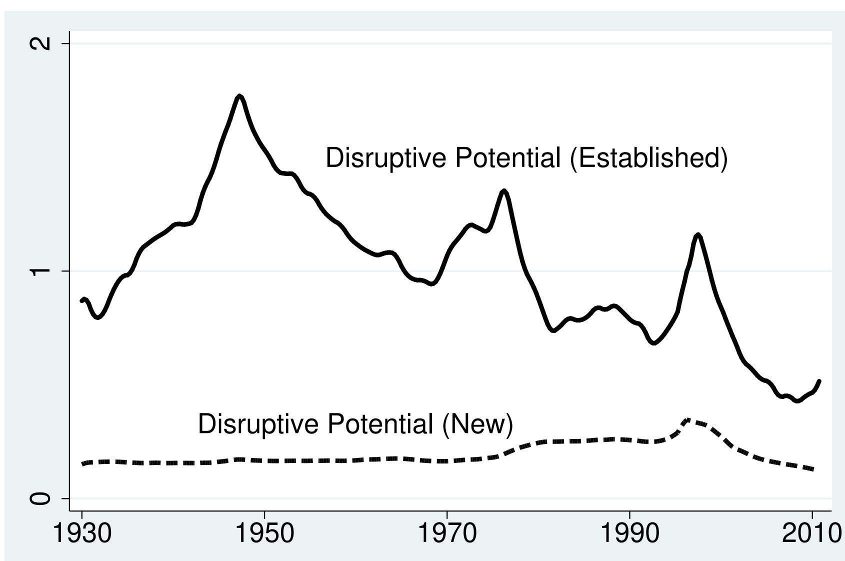
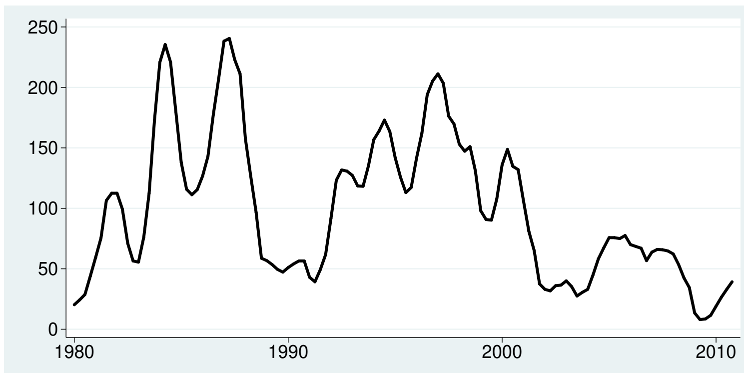


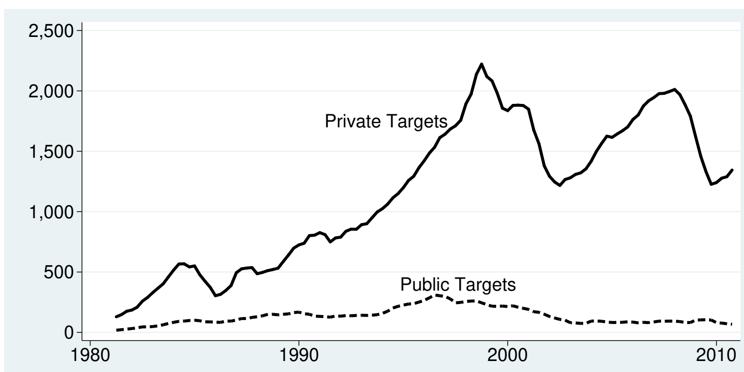
Figure III: Trends in Aggregate IPOs and Acquisitions

This figure reports the evolution between 1980 and 2010 the quarterly number of IPOs and acquisitions. For comparison, we also report disruptive potential from Figure II for the same period. 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. For comparison, we include *Disruptive Potential* (from Figure II) over the same period in the right panel. All series are reported as four quarter moving averages.

Panel A: IPO Volume



Panel B: Acquisition Volume



Panel C: Disruptive Potential

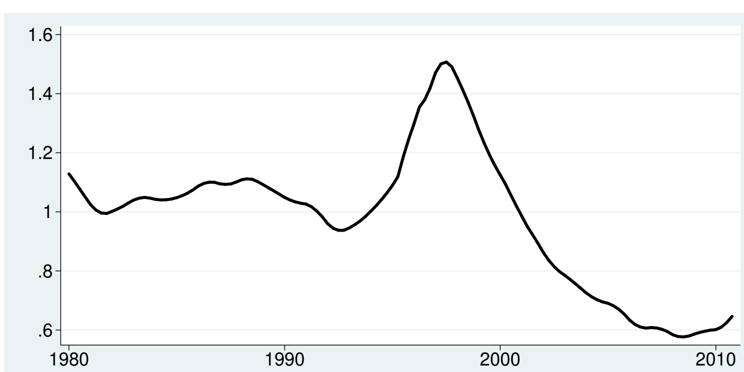
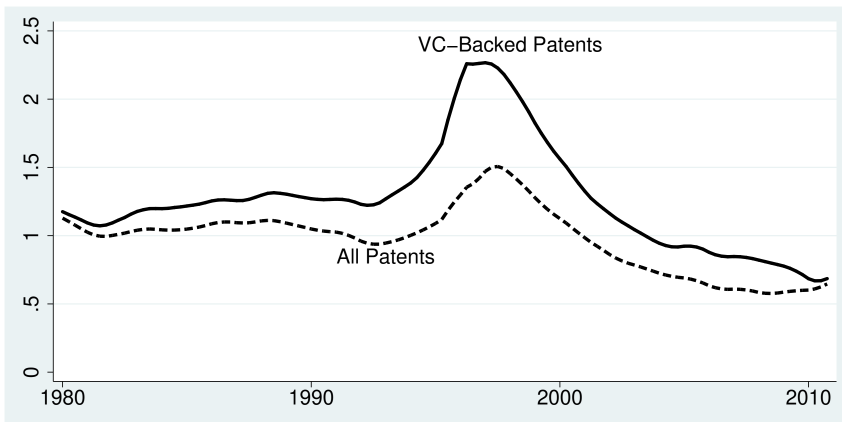


Figure IV: VC-Backed Startups: Disruptive Potential and Exit Trends

This figure compares trends among the startup sample to aggregate data. Panel A reports *Disruptive Potential* of patents held by VC-backed startups from 1980 to 2010 (solid line) and of all patents (dashed line). *Disruptive Potential* is defined at the patent level in Section III. The time series are constructed from the patent-level data as in Figure II. Panel B reports in the solid lines the percentage of startups that exit in the sample during the year via IPO or sell-out (left axis). The dashed lines report aggregate trends on IPOs and sell-outs of private targets and are reported in dashed lines as a fraction of lagged real GDP (right axis). Real GDP is in units of $100m. Aggregate data on IPOs and sell-outs are described in Figure III, except that in Panel B below, only private targets are included for sell-outs. All series are reported as four quarter moving averages.

Panel A: Disruptive Potential – Startup Sample vs. Aggregate Data



Panel B: Exits – Startup Sample vs. Aggregate Data

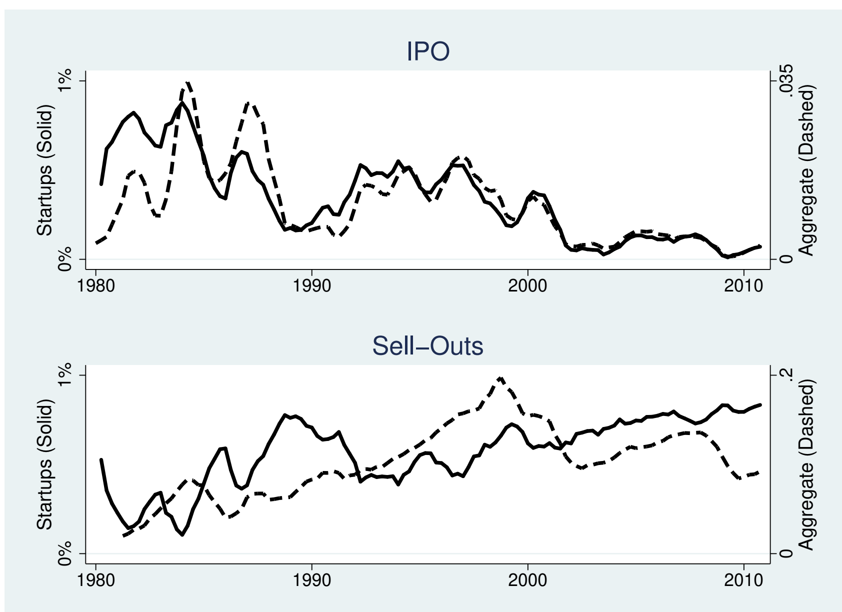


Table I: Changes in Patent Word Usage: Examples

This table reports, for five illustrative years between 1930 and 2010, 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 summary statistics for patent applications between 1930 and 2010 in Panel A and for the quarterly sample of venture-backed startups between 1980 and 2010 in Panel B. Startups are in the sample from their founding date until the quarter of their final outcome. Note that some startups remain private at the end of the sample period. The startup sample is further detailed in Section V.D. *Disruptive Potential* is defined at the patent level in Section III.B and at the startup-quarter level in Section III.D. Remaining variables are defined in Table A1. P25 and P75 denote the 25th and 75th percentiles. The underlying patent level measures are winsorized at the 1/99% level annually.

Panel A: Patent sample

	N	Mean	SD	P25	Median	P75
Disruptive Potential	6,594,248	1.64	1.81	0.51	1.27	2.34
Tech Breadth	6,594,143	0.42	0.22	0.24	0.47	0.60
Private Similarity	6,594,248	0.15	0.05	0.12	0.15	0.18
LI Similarity	6,594,248	0.11	0.05	0.06	0.09	0.13
Foreign Similarity	6,594,248	0.15	0.06	0.11	0.14	0.19
Originality	5,335,987	0.40	0.33	0.00	0.46	0.67
# of Cites	6,595,226	1.58	2.91	0.00	1.00	2.00
KPSS Value	1,781,386	9.75	23.69	0.73	3.25	9.16
mCD	4,245,716	0.56	1.83	0.00	0.00	0.38

Panel B: Startup-quarter sample

	N	Mean	SD	P25	Median	P75
Disruptive Potential	347,929	0.66	1.14	0.00	0.00	0.98
Tech Breadth	347,929	0.13	0.18	0.00	0.01	0.27
Private Similarity	347,929	0.06	0.06	0.00	0.06	0.11
LI Similarity	347,929	0.04	0.05	0.00	0.03	0.07
Foreign Similarity	347,929	0.05	0.06	0.00	0.04	0.09
Log(1+Firm Age)	347,929	3.07	1.15	2.40	3.18	3.76
No PatApps[q-1,q-20]	347,929	0.47	0.50	0.00	0.00	1.00
Log(1+PatApps[q-1,q-20])	347,929	0.79	0.97	0.00	0.69	1.39
Log(MTB) (q-2)	347,929	0.15	0.08	0.11	0.15	0.19
MKT Return [q-2,q-1]	347,929	0.01	0.13	-0.08	0.02	0.09
Q4	347,929	0.25	0.43	0.00	0.00	0.00
Originality	347,929	0.16	0.20	0.00	0.00	0.31
Log(1+Cites)	347,929	0.54	0.70	0.00	0.00	1.02
IPO rate (x100)	347,929	0.42	6.43	0.00	0.00	0.00
Sell-Out rate (x100)	347,929	0.73	8.50	0.00	0.00	0.00

Table III: Technological Disruptive Potential: Citation Patterns and Economic Value

This table presents regressions on a sample of patents granted between 1930 and 2010. Independent variables are defined in Table A1, while dependent variables are defined in Section III. To facilitate interpretation, all controls are standardized. All measures are winsorized at the 1/99% level annually. *LI Similarity* and *Foreign Similarity* are orthogonalized relative to *Private Similarity.* Fixed effects are included based on a patent's grant year (cohort) and technology category. Adjusted R^2 is reported as a percentage. Standard errors are clustered by the patent's grant year and are reported in parentheses. The symbols ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Log(1+Cites)	mCD	Log(1+KPSS Value)
	(1)	(2)	(3)
Panel A: Cohort fixed effects, no controls			
Disruptive Potential	0.035***	0.104***	0.080***
	(6.92)	(4.35)	(7.48)
Observations	6,033,343	4,245,630	1,697,481
R2 (%)	10.0	27.4	6.4
Panel B: Cohort-by-technology fixed effects, no controls			
Disruptive Potential	0.033***	0.078***	0.054***
	(8.93)	(4.62)	(5.99)
Observations	6,033,292	4,245,585	1,697,455
R2 (%)	15.9	30.0	12.1
Panel C: Cohort-by-technology fixed effects, with controls			
Disruptive Potential	0.030***	0.070***	0.033***
	(8.88)	(4.40)	(4.08)
Tech Breadth	-0.021***	-0.031***	-0.028**
	(-6.30)	(-4.17)	(-2.38)
Private Similarity	0.106***	0.064**	0.087***
	(15.87)	(2.29)	(6.53)
LI Similarity	0.128***	0.061*	0.363***
	(16.85)	(1.79)	(12.00)
Foreign Similarity	-0.064***	0.022	-0.534***
	(-10.64)	(0.85)	(-11.33)
Observations	6,033,230	4,245,584	1,697,435
R2 (%)	17.2	30.1	18.9

Table IV: Technological Disruptive Potential: Examples of Important Patents

This table reports the percentiles of various patent-level characteristics for important patents. Percentiles are cohort-adjusted, i.e., we remove year fixed effects before computing percentiles. In Panel A, we consider twelve unambiguous breakthrough patents. Patents before 1960 are from the USPTO's "Significant Historical Patents of the United States" and more recent patents are noted for the revenue they generated. In Panel B, we report summary statistics for the percentiles of a more comprehensive list of 101 patents between 1930 and 2010 identified by Kelly, Papanikolaou, Seru, and Taddy (2019) (henceforth, KPST) based on several on-line lists of "important" patents. These patents are listed in detail in Table A2. The percentiles for the KPST measure are taken directly from their Table A.6. All other variables are defined in Table A1. "Brdth" and "Orig" are short for *Tech Breadth* and *Originality*, respectively. The underlying patent-level measures are winsorized at 1/99% level annually.

Ex ante measureable:		Yes	No	Yes	No	No	Yes	Yes	
Patent	Year	DP	Cites	KPSS	mCD	KPST	Brth	Orig	Note
Panel A: Examples of Important Patents									
1,773,980	1930	0.88	0.7			0.98	0.64		TV
1,848,389	1932	0.73	0.68			0.94	0.3		Helicopter
2,404,334	1946	0.59	0.97			0.23	0.8		Jet Engine
2,524,035	1950	0.68	0.96	0.85		0.75	0.45	0.89	Transistor
2,569,347	1951	0.72	0.96	0.79		0.63	0.48	0.72	Junction Transistor
2,668,661	1954	1	0.87	0.8		0.98	0.85	0.83	Modern digital computer
2,835,548	1958	0.75	0.79			0.85	1	0.97	Satellite
2,929,922	1960	0.91	0.97	0.9		0.89	0.61		Laser
4,237,224	1980	0.96	0.98		0.99	1	0.62		Cohen/Boyer patent
4,399,216	1983	1	0.99		0.99	1	0.68	0.26	"Axel" patent
4,681,893	1987	0.7	1	0.58	0.99	N/A	0.3	0.4	Lipitor patent
6,285,999	2001	0.92	1		0.98	0.99	0.13	0.74	PageRank (Google)
Panel B: Summary of Important Patents Listed in Table A2									
Average:		0.71	0.75	0.68	0.61	0.84	0.53	0.55	
Median:		0.81	0.80	0.75	0.81	0.90	0.54	0.62	
Std error:		(0.03)	(0.02)	(0.04)	(0.06)	(0.02)	(0.03)	(0.04)	

Table V: Startup-Level Validity Tests: Perceived Disruption Risk

This table presents validity tests based on textual analysis of the 10-Ks of public peers of VC-backed startups. The sample is a cross-section of startups measured during the year in which they receive their first round of funding. As discussed in Section IV.C, 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.C. All variables are defined in Table A1 and the underlying patent-level measures are winsorized at the 1/99% level annually. 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:	Technology and change			Technology and competition			Technology, competition, and disruption		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Disruptive Potential	0.016***	0.012***	0.007**	0.039***	0.031***	0.025***	0.010***	0.007***	0.005**
	(4.75)	(3.89)	(2.24)	(4.49)	(3.64)	(2.82)	(4.48)	(3.42)	(2.15)
Constant	0.370***			1.401***			0.251***		
	(112.53)			(151.99)			(110.54)		
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
Observations	5,417	5,417	5,313	5,417	5,417	5,313	5,417	5,417	5,313
R2 (%)	0.4	6.6	30.2	0.3	4.9	21.0	0.4	8.1	25.3

Table VI: The Determinants of Startups' Exits - Baseline

This table presents cross-sectional tests relating startups' ex ante technological traits to their exit. The outcomes we consider are IPO and sell-out (acquisition). The sample is a quarterly panel of VC-backed startups from 1980-2010 and is described in Section III.D. 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 Table A1. *LI Similarity* and *Foreign Similarity* are orthogonalized relative to *Private Similarity*. The underlying patent-level measures 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. Adjusted R^2 is reported as a percentage. Standard errors are clustered by startup and are reported in parentheses. The symbols ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Competing Risk Hazard		OLS	
	IPO (1)	Sell-Out (2)	IPO (3)	Sell-Out (4)
Disruptive Potential	0.252***	-0.188***	0.082***	-0.067***
	(13.09)	(-7.55)	(4.04)	(-3.42)
Tech Breadth	0.510***	-0.210***	0.091***	-0.140***
	(10.02)	(-5.50)	(3.28)	(-4.39)
Private Similarity	0.127*	-0.479***	0.037	-0.433***
	(1.70)	(-8.62)	(1.06)	(-9.47)
LI Similarity	0.248***	-0.005	0.056*	-0.059
	(4.58)	(-0.12)	(1.80)	(-1.64)
Foreign Similarity	-0.056	-0.003	-0.017	0.059**
	(-1.42)	(-0.11)	(-0.93)	(2.39)
No PatApps[q-1,q-20]	1.648***	-2.172***	0.350***	-1.657***
	(10.75)	(-22.17)	(5.11)	(-15.09)
Log(1+PatApps[q-1,q-20])	0.327***	-0.056**	0.175***	-0.087***
	(9.65)	(-2.11)	(7.25)	(-2.97)
Log(MTB) (q-2)	0.133***	0.162***	0.151***	0.044
	(5.15)	(8.94)	(4.20)	(0.82)
MKT Return [q-2,q-1]	0.341***	0.004	0.046***	0.036*
	(11.92)	(0.16)	(3.55)	(1.79)
Q4	-0.059	0.114**	0.137***	0.370***
	(-0.77)	(2.05)	(2.86)	(5.75)
Originality	-0.125***	-0.175***	-0.028	-0.178***
	(-3.11)	(-6.01)	(-1.44)	(-7.34)
Log(1+Cites)	0.171***	0.118***	0.075***	0.152***
	(3.95)	(3.96)	(3.64)	(5.38)
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	346,490	345,403	342,146	342,146
R2 (%)	N/A	N/A	0.5	0.6

Table VII: The Determinants of Startups' Exits - Decomposition

This table presents cross-sectional tests relating startups' ex ante technological traits to their exit. Each of the models repeats the corresponding model from Table VI, but replaces the main variable *Disruptive Potential* with a decomposition by focusing on a subset of words in the patent corpus for each year. *Disruptive Potential (Established)* is defined in Equation 5 based on words that are at least ten years old in a given year. *Disruptive Potential (New)* is defined in Equation 6 based on words that less than ten years old in a given year. For brevity, we only report the coefficients on the decomposed variables. To facilitate interpretation, coefficients for OLS estimates report the incremental % change in a given outcome, and *Disruptive Potential (Established)* and *Disruptive Potential (New)* are standardized. Independent variables are lagged one quarter unless explicitly noted. The underlying patent-level measures are winsorized at the 1/99% level annually. Adjusted R^2 is reported as a percentage. Standard errors are clustered by startup and are reported in parentheses. The symbols ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Competing Risk Hazard		OLS	
	IPO (1)	Sell-Out (2)	IPO (3)	Sell-Out (4)
Disruptive Potential (Established)	0.306***	-0.146***	0.116***	-0.038*
	(10.20)	(-6.47)	(5.54)	(-1.78)
Disruptive Potential (New)	0.022	-0.066***	-0.014	-0.036**
	(0.99)	(-2.90)	(-0.90)	(-1.97)
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	346,490	345,403	342,146	342,146
R2 (%)	N/A	N/A	0.5	0.6

Table VIII: Startup-Level Validity Tests: IPO Subsample

This table presents validity tests based on a subsample of VC-backed startups that go public. *Primary Share %* is the fraction of IPO shares that are newly issued (as opposed to existing shares sold by insiders) and based on data from SDC. *No Secondary Shares* equals one when all shares are newly issued and zero if SDC records any sales by insiders. Columns (3)-(5) examine a subsample after 1997 where we are able to merge in both public firm identifiers (GVKEY) and obtain data on the product space of the firm (as of the startup's 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). *Disruptive Potential* is measured in the quarter preceding the IPO. 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.

	Primary Share % (1)	No Secondary Shares (2)	HHI (3)	Log(TSimm) (4)	Product Mkt Fluidity (5)
Disruptive Potential	0.152***	0.042***	-0.012**	0.073**	0.245**
	(4.78)	(4.72)	(-2.18)	(2.10)	(2.16)
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	1,167	1,167	523	523	524
R2 (%)	8.4	9.1	1.7	1.7	2.5

Table IX: Determinants of Startups' Exits - Financing

This table presents cross-sectional tests relating startups' ex ante technological traits to their exit. Each of the models repeats the corresponding model from Table VI, 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 *Disruptive Potential*. To facilitate interpretation, coefficients for OLS estimates report the incremental % change in a given outcome, and *Disruptive Potential* is standardized. Independent variables are lagged one quarter unless explicitly noted. The underlying patent-level measures are winsorized at the 1/99% level annually. Adjusted R^2 is reported as a percentage. Standard errors are clustered by startup 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)
Disruptive Potential	0.263***	-0.170***	0.075***	-0.077***
	(13.66)	(-6.82)	(3.70)	(-4.00)
log(CumVCFunding)	0.074***	0.138***	0.059***	0.119***
	(4.56)	(10.18)	(8.83)	(15.01)
No Funding[q-1,q-20]	-0.927***	-2.508***	-0.126**	0.130**
	(-5.35)	(-9.79)	(-2.22)	(1.99)
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	346,490	345,403	342,146	342,146
R2 (%)	N/A	N/A	0.7	0.8

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

This table presents dynamic cross-sectional tests relating startups' ex ante technological traits to their exit over several horizons. In Panel A, column 1 repeats the OLS model examining IPO exits from column 3 in Table VI. 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 VI. Independent variables are 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 startup 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						
Disruptive Potential	0.082***	0.299***	0.420***	0.424**	0.472**	0.466**
	(4.04)	(3.88)	(3.14)	(2.49)	(2.34)	(2.07)
Tech Breadth	0.091***	0.362***	0.714***	0.973***	1.167***	1.229***
	(3.28)	(3.54)	(3.85)	(3.90)	(3.81)	(3.52)
Private Similarity	0.037	0.120	0.215	0.261	0.492	0.623
	(1.06)	(0.94)	(0.95)	(0.87)	(1.34)	(1.52)
LI Similarity	0.056*	0.182	0.287	0.265	0.313	0.270
	(1.80)	(1.57)	(1.40)	(0.95)	(0.91)	(0.69)
Foreign Similarity	-0.017	-0.045	-0.029	0.007	-0.045	-0.089
	(-0.93)	(-0.64)	(-0.22)	(0.04)	(-0.21)	(-0.36)
Panel B: Exit by Sell-Out						
Disruptive Potential	-0.067***	-0.191**	-0.145	-0.055	0.011	0.089
	(-3.42)	(-2.46)	(-0.92)	(-0.25)	(0.04)	(0.28)
Tech Breadth	-0.140***	-0.600***	-1.086***	-1.501***	-1.899***	-2.056***
	(-4.39)	(-4.85)	(-4.60)	(-4.43)	(-4.43)	(-4.10)
Private Similarity	-0.433***	-1.468***	-2.123***	-2.052***	-1.485**	-0.628
	(-9.47)	(-8.23)	(-6.24)	(-4.28)	(-2.53)	(-0.93)
LI Similarity	-0.059	-0.164	-0.012	0.303	0.652	1.297**
	(-1.64)	(-1.14)	(-0.04)	(0.75)	(1.29)	(2.20)
Foreign Similarity	0.059**	0.158	0.059	-0.219	-0.460	-0.924**
	(2.39)	(1.60)	(0.31)	(-0.79)	(-1.30)	(-2.24)
Panel C: Still Private						
Disruptive Potential	-0.056*	-0.267**	-0.564***	-0.745***	-0.859***	-0.880**
	(-1.94)	(-2.41)	(-2.77)	(-2.76)	(-2.70)	(-2.54)
Tech Breadth	0.044	0.147	0.175	0.223	0.402	0.516
	(0.99)	(0.88)	(0.57)	(0.53)	(0.79)	(0.89)
Private Similarity	0.359***	1.301***	2.049***	2.235***	1.778***	1.305*
	(5.98)	(5.82)	(5.01)	(4.04)	(2.69)	(1.77)
LI Similarity	0.045	0.137	0.092	-0.038	-0.311	-0.842
	(0.90)	(0.73)	(0.26)	(-0.08)	(-0.54)	(-1.30)
Foreign Similarity	-0.088***	-0.279**	-0.409*	-0.390	-0.392	-0.161
	(-2.65)	(-2.20)	(-1.73)	(-1.19)	(-0.98)	(-0.36)

Table XI: Explaining Aggregate IPO and Sell-Out Rates

This table presents the out-of-sample performance of predictive models of startups' exit using variables standard in the IPO and acquisition literature (the "Base" model) and a model which augments the "Base" model with the new text-based technological variables (the "Text" model). Panel A examines IPO exits and Panel B examines sell-outs. In a given test (column 1), we estimate a Fama and MacBeth (1973) regression quarter-by-quarter where the dependent variable is a dummy indicating an IPO exit (Panel A) or indicating a sell-out exit (Panel B) based on the horizon listed in column 2 (ranging from one quarter to three years) and using the ex ante measurable independent variables in Table VI. This model is fitted using the early part of our sample, which begins in 1980 and ends before the out-of-sample period (noted in column 3). These fitted Fama-MacBeth coefficients from the early period are then used in the out-of-sample post period (listed in column 3) to predict the average IPO rate and sell-out rate. These predicted exit rates are then compared to the actual rates to compute the fraction of the disappearing IPOs or surging sell-outs anomaly that is explained by either the "Base" model or the "Text" model as noted in columns (5) to (8). Column 9 reports the percentage of each anomaly that cannot be explained by the base model that is explained by the Text model. All probabilities in columns (4)-(8) are reported as percentage points.

Test	Prediction Horizon	Post Period	True Exit Rate	Predicted Exit Rate		Error		Text Impr
				Base	Text	Base	Text	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: IPO Exits								
1	1Q	[1996,2010]	0.33	0.84	0.75	0.50	0.42	**16%**
2	1Q	[1998,2010]	0.27	0.85	0.76	0.58	0.49	**16%**
3	1Q	[2000,2010]	0.22	0.85	0.75	0.63	0.53	**16%**
4	1Y	[1996,2010]	1.27	3.34	2.89	2.07	1.63	**21%**
5	2Y	[1996,2010]	2.45	6.50	5.59	4.05	3.15	**22%**
6	3Y	[1996,2010]	3.56	9.46	8.17	5.90	4.60	**22%**
Panel B: Sell-Out Exits								
7	1Q	[1996,2010]	0.86	0.60	0.75	-0.25	-0.11	**57%**
8	1Q	[1998,2010]	0.91	0.63	0.80	-0.27	-0.10	**62%**
9	1Q	[2000,2010]	0.95	0.68	0.87	-0.27	-0.08	**71%**
10	1Y	[1996,2010]	3.49	2.52	2.96	-0.98	-0.54	**45%**
11	2Y	[1996,2010]	7.19	5.30	5.94	-1.88	-1.25	**34%**
12	3Y	[1996,2010]	10.96	8.21	8.93	-2.75	-2.03	**26%**

Table XII: Explaining Aggregate IPO Rates (Small vs Big IPOs)

This table presents the out-of-sample performance of predictive models of startups' exit using variables standard in the IPO and acquisition literature (the "Base" model) and a model which augments the "Base" model with the new text-based technological variables (the "Text" model). Panel A examines small IPO exits and Panel B big IPO exits. We define an IPO as "small" if its pre-IPO sales are below the median in our sample ($25 million), and as "large" if its pre-IPO sales exceeds that amount. In a given test (column 1), we estimate a Fama and MacBeth (1973) regression quarter-by-quarter where the dependent variable is a dummy indicating an IPO exit (Panel A) or indicating a sell-out exit (Panel B) based on the horizon listed in column 2 (ranging from one quarter to three years) and using the ex ante measurable independent variables in Table VI. This model is fitted using the early part of our sample, which begins in 1980 and ends before the out-of-sample period (noted in column 3). These fitted Fama-MacBeth coefficients from the early period are then used in the out-of-sample post period (listed in column 3) to predict the average IPO rate and sell-out rate. These predicted exit rates are then compared to the actual rates to compute the fraction of the disappearing IPOs or surging sell-outs anomaly that is explained by either the "Base" model or the "Text" model as noted in columns (5) to (8). Column 9 reports the percentage of each anomaly that cannot be explained by the base model that is explained by the Text model. All probabilities in columns (4)-(8) are reported as percentage points.

Test (1)	Prediction Horizon (2)	Post Period (3)	True Exit Rate (4)	Predicted Exit Rate Base (5)	Predicted Exit Rate Text (6)	Error Base (7)	Error Text (8)	Text Impr (9)
Panel A: Small IPO Exits								
7	1Q	[1996,2010]	0.15	0.35	0.28	0.21	0.13	**36%**
8	1Q	[1998,2010]	0.12	0.38	0.29	0.26	0.18	**32%**
9	1Q	[2000,2010]	0.08	0.38	0.30	0.30	0.21	**28%**
10	1Y	[1996,2010]	0.55	1.45	1.02	0.90	0.47	**48%**
11	2Y	[1996,2010]	1.06	2.86	2.10	1.80	1.04	**42%**
12	3Y	[1996,2010]	1.53	4.17	3.22	2.65	1.69	**36%**
Panel B: Big IPO Exits								
1	1Q	[1996,2010]	0.13	0.35	0.38	0.22	0.25	**-13%**
2	1Q	[1998,2010]	0.11	0.35	0.38	0.24	0.27	**-11%**
3	1Q	[2000,2010]	0.10	0.34	0.36	0.24	0.27	**-9%**
4	1Y	[1996,2010]	0.51	1.36	1.41	0.86	0.91	**-6%**
5	2Y	[1996,2010]	0.99	2.64	2.54	1.65	1.55	**6%**
6	3Y	[1996,2010]	1.46	3.86	3.59	2.39	2.12	**11%**

Table XIII: Explaining Aggregate IPO and Sell-Out Rates (Stable vs Fluid Markets)

This table presents the out-of-sample performance of predictive models of startups' exit with variables standard in the IPO and acquisition literature (the "Base" model) and a model which augments the "Base" model with the new text-based technological variables (the "Text" model). Panel A examines IPO exits and Panel B examines sell-outs. The procedure is analogous to that described in Table XI, except each test is repeated for two subsamples: *Stable Markets* and *Fluid Markets*, which are defined in Section VII.C. We omit the model-implied out-of-sample probabilities to conserve space.

			Stable Market Subsample				**Fluid Market Subsample**			
Test (1)	Pred. Horizon (2)	Post Period (3)	True Rate (4)	Base Error (5)	Text Error (6)	Text Impr (7)	True Rate (8)	Base Error (9)	Text Error (10)	Text Impr (11)
			Panel A: IPO Exits							
1	1Q	[1996,2010]	0.30	0.41	0.31	**26%**	0.37	0.62	0.67	**-7%**
2	1Q	[1998,2010]	0.26	0.45	0.33	**27%**	0.29	0.75	0.79	**-5%**
3	1Q	[2000,2010]	0.24	0.44	0.32	**29%**	0.20	0.86	0.89	**-4%**
4	1Y	[1996,2010]	1.14	1.69	1.29	**24%**	1.42	2.57	2.29	**11%**
5	2Y	[1996,2010]	2.20	3.32	2.60	**22%**	2.74	5.02	4.27	**15%**
6	3Y	[1996,2010]	3.23	4.75	3.68	**22%**	3.96	7.38	6.51	**12%**
			Panel B: Sell-Out Exits							
7	1Q	[1996,2010]	0.85	-0.16	-0.04	**73%**	0.86	-0.37	-0.21	**44%**
8	1Q	[1998,2010]	0.89	-0.17	-0.04	**75%**	0.93	-0.40	-0.19	**52%**
9	1Q	[2000,2010]	0.91	-0.16	-0.01	**93%**	0.98	-0.40	-0.17	**57%**
10	1Y	[1996,2010]	3.43	-0.59	-0.34	**43%**	3.55	-1.46	-0.79	**46%**
11	2Y	[1996,2010]	6.97	-1.05	-0.89	**15%**	7.39	-2.90	-1.56	**46%**
12	3Y	[1996,2010]	10.52	-1.44	-1.43	**1%**	11.38	-4.32	-2.32	**46%**

Table A1: Variable Definitions

Patent-Level Variables	
Disruptive Potential	See Equation 2 and Section III.B.
Tech Breadth	See Equation 3 and Section III.C.
LI Similarity	See Equation 4 and Section III.C.
Private Similarity	Similar to *LI Similarity*. See Section III.C.
Foreign Similarity	Similar to *LI Similarity*. See Section III.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.
mCD	From Funk and Owen-Smith (2016).
Originality	The originality of a focal patent is defined as 1 minus the HHI of the technology fields of the patents cited by the focal patent (Trajtenberg, Henderson, and Jaffe (1997)). We use the adjustment given in Hall, Jaffe, and Trajtenberg (2001) to reduce bias for patents that contain few backward citations. We convert U.S. Patent Classifications to the NBER technology codes so that *Tech Breadth* and *Originality* are based on the same granularity of technology classifications.
Disruptive Potential (Established)	See Equation 5 and Section III.B.
Disruptive Potential (New)	See Equation 6 and Section III.B.
Startup-Quarter Variables	
Disruptive Potential	The depreciated sum of patent-level *Disruptive Potential* 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 startup applied for. See Section III.D for more.
Tech Breadth	Converted to startup-quarter like *Disruptive Potential*.
Private Similarity	Converted to startup-quarter like *Disruptive Potential*.
LI Similarity	Converted to startup-quarter like *Disruptive Potential*.
Foreign Similarity	Converted to startup-quarter like *Disruptive Potential*.
Log(1+Cites)	Log of the stock of citations. Citations for a startup-quarter is the sum of the *# of Cites* (patent-level variable defined above) for patents the startup applies for in the quarter. Note that this is forward-looking. The stock is computed using a quarterly depreciation of 5%.
Originality	Converted to startup-quarter like *Disruptive Potential*.
No PatApps[q-1,q-20]	Dummy variable equal to one if the startup has not applied for a patent (which was eventually granted) during the last 20 quarters.
Log(1+PatApps[q-1,q-20])	The # of (granted) patent applications in the last 20 quarters.
IPO	One if the startup goes public in the quarter, zero before.
Sell-out	One if the startup is acquired in the quarter, zero before.
Disruptive Potential (Established)	Converted to startup-quarter like *Disruptive Potential*.
Disruptive Potential (New)	Converted to startup-quarter like *Disruptive Potential*.
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]	From Ken French's daily factor file using geometric compounding.
Q4	One if $t-1$ is the fourth quarter (and t is the first quarter), else zero.

Table A2: Percentiles of various statistics for a sample of important patents
The patents below are the 1930-2010 subset of key important patents listed in Kelly, Papanikolaou, Seru, and Taddy (2019) (henceforth, KPST) over which the textual measures in this paper are defined. The percentiles for the KPST measure are taken directly from their Table A.6. Remaining variables are defined in Table A1. "Brdth" and "Orig" are short for *Tech Breadth* and *Originality*, respectively. The underlying patent-level measures are winsorized at 1/99% level annually. Percentiles are cohort-adjusted, i.e., we remove year fixed effects before computing percentiles.

Ex ante measureable:		Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Patent	Grant Year	Dsrpt Potent	Cites	KPSS	mCD	KPST	Brdth	Orig	Priv Simm	LI Simm	Frgn Simm
Panel A: Summary statistics of percentiles in Panel B											
Average:		0.71	0.75	0.68	0.61	0.84	0.53	0.55	0.52	0.63	0.57
Median:		0.81	0.80	0.75	0.81	0.90	0.54	0.62	0.51	0.69	0.59
Std error:		(0.03)	(0.02)	(0.04)	(0.06)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Panel B: Percentiles of various measures for breakthrough patents											
1,773,079	1930	0.10	0.70			0.95	0.69		0.91	0.68	0.85
1,773,080	1930	0.50	0.70			0.95	0.31		0.18	0.28	0.37
1,773,980	1930	0.88	0.70			0.98	0.64		0.49	0.97	0.91
1,800,156	1931	0.86	0.70			0.97	0.47		0.83	0.59	0.86
1,821,525	1931	0.48	0.68			0.55	0.98		0.96	0.96	0.92
1,835,031	1931	0.83	0.68	0.75		0.90	0.96		0.60	0.96	0.80
1,848,389	1932	0.73	0.68			0.94	0.30		0.94	0.79	0.89
1,867,377	1932	0.12	0.69			0.75	0.36		0.52	0.54	0.35
1,925,554	1933	0.88	0.68			0.92	0.83		0.20	0.72	0.59
1,929,453	1933	0.95	0.66			0.98	0.26		0.82	0.70	0.90
1,941,066	1933	0.78	0.68			0.93	0.74		0.19	0.97	0.86
1,948,384	1934	0.78	0.66			0.87	0.79		0.21	0.76	0.66
1,949,446	1934	0.50	0.66			0.55	0.31		0.83	0.57	0.48
1,980,972	1934	1.00	0.65			0.98	0.18		0.84	0.79	1.00
2,021,907	1935	0.60	0.67			0.89	0.99		0.28	0.82	0.65
2,059,884	1936	0.94	0.66	0.63		0.59	0.46		0.41	0.72	0.89
2,071,250	1937	0.98	0.67	0.84		0.89	0.30		0.97	0.91	1.00
2,087,683	1937	0.86	0.66			0.92	0.68		0.37	0.84	0.91
2,153,729	1939	1.00	0.65			0.96	0.16		0.67	0.73	1.00
2,188,396	1940	0.78	0.63	0.60		1.00	0.17		0.83	0.62	0.80
2,206,634	1940	0.93	0.65			0.98	0.57		0.68	0.71	0.97
2,230,654	1941	0.93	0.62			0.93	0.25		0.88	0.75	0.51
2,258,841	1941	0.38	0.56			0.23	0.88		0.72	0.52	0.83
2,292,387	1942	0.52	0.56			0.95	0.94		0.58	0.81	0.63
2,297,691	1942	0.14	0.62			0.62	0.87		0.67	0.69	0.76
2,329,074	1943	0.91	0.89			0.56	0.20		0.99	0.95	0.98
2,390,636	1945	0.18	0.95			0.79	0.57		0.08	0.14	0.33
2,404,334	1946	0.59	0.97			0.23	0.80		0.83	0.79	0.96
2,436,265	1948	0.37	0.95			0.74	0.39	0.26	0.76	0.75	0.66
2,451,804	1948	0.92	0.93			0.74	0.20	0.06	0.94	0.80	0.50
2,495,429	1950	0.85	0.96			0.21	0.86	0.92	0.39	0.71	0.32

Continued on next page

Ex ante measureable:		Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Patent	Grant Year	Dsrpt Potent	Cites	KPSS	mCD	KPST	Brdth	Orig	Priv Simm	LI Simm	Frgn Simm
2,524,035	1950	0.68	0.96	0.85		0.75	0.45	0.89	0.66	0.93	0.90
2,543,181	1951	0.74	0.96			0.63	0.32	0.61	0.93	0.85	0.71
2,569,347	1951	0.72	0.96	0.79		0.63	0.48	0.72	0.79	0.99	0.95
2,642,679	1953	0.10	0.81			0.55	0.67	0.40	0.97	0.68	0.92
2,668,661	1954	1.00	0.87	0.80		0.98	0.85	0.83	0.01	0.01	0.01
2,682,050	1954	0.84	0.42			0.77	0.99	0.18	0.29	0.66	0.46
2,682,235	1954	0.88	0.63			0.60	0.99		0.10	0.23	0.15
2,691,028	1954	0.90	0.40			0.96	0.05		0.97	0.92	0.94
2,699,054	1955	0.21	0.97			0.97	0.23	0.77	0.98	0.97	0.98
2,708,656	1955	1.00	0.97			0.99	0.93		0.01	0.01	0.01
2,708,722	1955	0.91	0.97			0.78	0.44	0.99	0.08	0.75	0.43
2,717,437	1955	0.52	0.80			0.43	0.45	0.17	0.03	0.06	0.11
2,724,711	1955	0.99	0.39			0.82	0.06		0.46	0.75	0.78
2,752,339	1956	0.79	0.91			0.88	0.04	0.20	0.95	0.95	0.96
2,756,226	1956	0.89	0.60			0.71	0.18	0.98	0.34	0.63	0.86
2,797,183	1957	1.00	0.80			0.90	0.24	0.98	0.84	0.70	0.87
2,816,721	1957	0.59	0.95			0.72	0.46		0.42	0.52	0.72
2,817,025	1957	0.24	0.97			0.71	0.48		0.90	1.00	0.77
2,835,548	1958	0.75	0.79			0.85	1.00	0.97	0.65	0.42	0.55
2,866,012	1958	0.25	0.98			0.81	0.54	0.15	0.97	0.97	0.87
2,879,439	1959	0.72	0.97			0.77	0.80	0.55	0.44	0.82	0.55
2,929,922	1960	0.91	0.97	0.90		0.89	0.61		0.39	0.66	0.58
2,937,186	1960	0.99	0.58			0.89	0.12	0.13	0.99	0.97	1.00
2,947,611	1960	0.47	0.34	0.58		0.77	0.57		0.94	0.95	0.94
2,956,114	1960	0.45	0.90	0.71		0.74	0.55	0.39	0.98	0.98	0.90
2,981,877	1961	0.82	0.98			0.98	0.42	0.10	0.60	0.81	0.47
3,057,356	1962	0.89	0.97			0.93	0.54	0.09	0.55	0.94	0.58
3,093,346	1963	0.92	0.98			0.93	0.82	0.82	0.40	0.53	0.51
3,097,366	1963	0.28	0.55			0.41	0.99	0.73	0.86	0.62	0.77
3,118,022	1964	0.33	0.29	0.89		0.70	0.71		0.70	0.77	0.60
3,156,523	1964	0.05	0.46			0.85	0.25	1.00	0.86	0.83	0.98
3,174,267	1965	0.42	0.89	0.48		0.55	0.94	0.73	0.45	0.09	0.16
3,220,816	1965	0.06	0.54			0.85	0.42	0.09	0.03	0.14	0.09
3,287,323	1966	0.29	0.32	0.56		0.70	0.13	0.39	0.56	0.98	0.99
3,478,216	1969	0.42	0.80			0.84	0.37	0.62	0.35	0.64	0.46
3,574,791	1971	0.21	0.96	0.93		0.82	0.19		0.79	1.00	0.99
3,663,762	1972	0.38	0.97	0.84		0.78	0.68	0.26	0.13	0.40	0.09
3,789,832	1974	0.76	0.89			0.74	0.91	0.79	0.54	0.85	0.65
3,858,232	1974	0.32	0.98	0.97		0.71	0.78	0.95	0.58	0.94	0.77
3,906,166	1975	0.86	0.92	0.55		0.71	0.93	0.29	0.45	0.73	0.42
4,136,359	1979	0.71	0.76		0.89	0.97	0.69		0.63	0.98	0.83
4,229,761	1980	0.64	0.30			0.92	0.80	1.00	0.02	0.17	0.09
4,237,224	1980	0.96	0.98		0.99	1.00	0.62		0.05	0.22	0.14
4,363,877	1998		1.00		0.98	1.00					
4,371,752	1983	0.96	0.99		0.95	0.94	0.30	0.28	0.72	0.99	0.82
4,399,216	1983	1.00	0.99		0.99	1.00	0.68	0.26	0.04	0.18	0.07
4,437,122	1993		1.00	0.61	0.99	1.00					
4,464,652	1984	0.39	0.99	0.90	0.95	0.89	0.86	0.78	0.95	0.82	0.80

Continued on next page

Ex ante measureable:		Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Patent	Grant Year	Dsrpt Potent	Cites	KPSS	mCD	KPST	Brdth	Orig	Priv Simm	LI Simm	Frgn Simm
4,468,464	1984	1.00	0.92		0.16	1.00	0.58		0.04	0.16	0.12
4,590,598	1986	0.67	0.28	0.93	0.19	0.58	0.78	0.91	0.82	0.93	0.84
4,634,665	1987	0.96	0.78		0.97	0.99	0.53		0.05	0.18	0.07
4,683,195	1987	0.93	0.11	0.42	0.99	0.97	0.72		0.34	0.63	0.55
4,683,202	1990		0.14	0.38	0.99	0.94					
4,736,866	1988	1.00	0.92		0.96	1.00	0.47	0.24	0.02	0.07	0.02
4,744,360	1988	0.80	0.83		0.21	0.91	0.68	0.18	0.55	0.64	0.57
4,799,258	1989	0.80	1.00		0.82	0.95	0.32	0.95	0.20	0.55	0.14
4,816,397	1989	1.00	0.94		0.48	0.98	0.60	0.99	0.20	0.41	0.23
4,816,567	1989	0.94	0.15	0.82	0.95	0.99	0.66	0.58	0.27	0.54	0.34
4,838,644	1989	0.78	0.88		0.85	0.92	0.95	0.20	0.82	0.93	0.87
4,889,818	1989	0.92	0.99	0.53	0.05	0.98	0.71	0.20	0.50	0.75	0.65
4,965,188	1990	0.93	0.99	0.49	0.02	0.98	0.75	0.20	0.40	0.72	0.61
5,061,620	1991	0.93	0.99		0.98	1.00	0.36		0.11	0.16	0.10
5,071,161	1991	0.48	1.00		0.94	0.67	0.94	0.20	0.68	0.47	0.39
5,108,388	1992	0.91	0.88	0.42	0.15	0.97	0.85	0.45	0.42	0.54	0.36
5,149,636	1992	0.66	0.48		0.24	0.99	0.52	0.19	0.05	0.18	0.07
5,179,017	1993	0.95	0.41		0.32	1.00	0.31		0.16	0.33	0.24
5,184,830	1993	0.62	0.94		0.31	0.98	0.34	0.43	0.84	0.88	0.89
5,194,299	1993	0.51	0.94	0.81	0.11	0.73	0.31	0.81	0.62	0.48	0.56
5,225,539	1993	0.98	0.99		0.99	1.00	0.52		0.10	0.26	0.17
5,272,628	1993	1.00	0.99	0.99	0.40	0.99	0.23	0.78	0.14	0.44	0.13
5,747,282	1998	1.00	0.45	0.08	0.95	0.97	0.15		0.39	0.30	0.32
5,770,429	1998	0.99	0.01		0.79	0.61	0.13	0.86	0.36	0.27	0.31
5,837,492	1998	1.00	0.01	0.04		0.83	0.28		0.39	0.39	0.40
5,939,598	1999	1.00	0.57	0.53	0.21	1.00	0.31	0.68	0.21	0.28	0.27
5,960,411	1999	0.87	1.00	1.00	0.22	1.00	0.03	0.85	0.18	0.50	0.09
6,230,409	2001	0.53	0.07		0.32	0.75	0.96	0.74	0.80	0.24	0.38
6,285,999	2001	0.92	1.00		0.98	0.99	0.13	0.74	0.13	0.47	0.12
6,331,415	2001	0.65	0.97	0.95	0.18	0.99	0.60	0.47	0.39	0.64	0.49
6,455,275	2002	0.99	0.45		0.04	0.98	0.13	0.12	0.45	0.40	0.46
6,574,628	2003	0.87	0.98		0.13	1.00	0.04	0.74	0.29	0.65	0.22
6,955,484	2005	0.07	0.75		0.79	0.78	0.68	0.63	0.63	0.24	0.35
6,985,922	2006	0.83	0.98		0.92	0.93	0.10	0.90	0.21	0.62	0.11

Internet Appendix for

“Technological Disruptive Potential and the Evolution of IPOs and Sell-Outs”

August 2019

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

IA.A Defining the Entity Type of Patents’ Assignees

To classify if a patent is granted to (A) a private, domestic U.S. firm, (B) an international firm, or (C) a U.S. public firm, we use the following procedure. First, we find all patents assigned to public firms. We obtain the GVKEY for assignees from the NBER patent dataset, and augment this with 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”.

IA.B Matching patents to VentureXpert

We download all data on firms receiving venture capital funding starting in 1970 and ending in 2013 from VentureXpert using SDC Platinum. In addition to the dates of venture financing, we also download data indicating each portfolio company’s founding date, 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. 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.[34] 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.[35] 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 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.[36] 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] Firms can receive patents before VC funding.

[35] http://www-personal.umich.edu/~nwasi/programs.html

[36] 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.

IA.C Additional Results

1. Figure IA1 presents the evolution of text-based characteristics of the aggregate patent corpus from 1930 to 2010. 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. We also 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.

2. Table IA1 presents information on the timing of key life events for startups in the main analysis sample.

3. Table IA2 presents robustness tests of the main results on the determinants of startups' exit from Table VI.

4. Table IA3 presents regressions of startups' financing on their technological characteristics.

5. Table IA4 presents subsample tests of the main OLS models on the determinants of startups' exit from Table VI. The subsamples are based on the date of the observation.

Figure IA1: 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 III. 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.

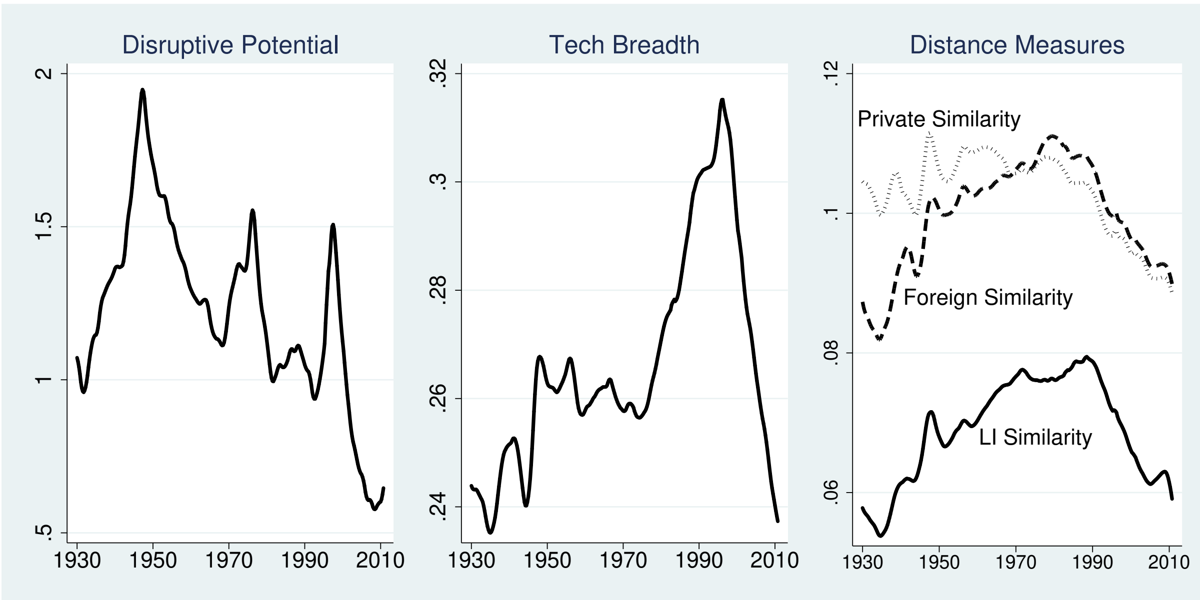


Table IA1: Years between keys events for ventured-backed Startups

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

Panel A: Events after the startup's founding

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

Panel B: Events after the startup's first patent

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

Table IA2: Robustness of baseline results

This table presents robustness tests of the main results in Table VI. For brevity, we only report the main coefficient on *Disruptive Potential* for each test. Each row corresponds to an alteration of the main test. Aside from the listed alteration, each of the models within a row repeats the corresponding model in the same column of Table VI. To facilitate interpretation, coefficients for OLS estimates report the incremental % change in a given outcome, and *Disruptive Potential* is standardized and lagged one quarter. The underlying patent-level measures are winsorized at the 1/99% level annually. Standard errors are clustered by startup unless otherwise noted 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
Test #	Test alteration	(1)	(2)	(3)	(4)
(1)	Cluster standard errors by technology	0.252***	-0.188***	0.081***	-0.067***
		(13.56)	(-8.90)	(8.25)	(-3.98)
(2)	Cluster standard errors by year	0.252***	-0.188***	0.081***	-0.067**
		(7.84)	(-5.78)	(3.13)	(-2.47)
(3)	Cluster standard errors by firm cohort	0.252***	-0.188***	0.081***	-0.067***
		(8.54)	(-5.97)	(3.40)	(-2.94)
(4)	Only controls: *No PatApps[q-1,q-20]*	0.267***	-0.098***	0.091***	-0.035*
	and *Log(1+PatApps[q-1,q-20])*	(15.03)	(-4.70)	(4.73)	(-1.94)
(5)	Exclude *Log(1+Cites)* as control	0.267***	-0.167***	0.093***	-0.045**
		(14.26)	(-6.98)	(4.59)	(-2.35)
(6)	Recode sell-outs as "liquidations" if		-0.183***		-0.053***
	exit value below $25m (2009 dollars)		(-6.76)		(-2.88)
(7)	Cross-section as of exit date	0.360***	-0.449***	2.041**	-2.041**
	for IPO and Sell-Out firms	(10.22)	(-7.69)	(2.41)	(-2.41)
		N=4,019	N=4,019	N=3,913	N=3,913

Table IA3: 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 (*Cum.Funds*) the firm receives between its founding and quarter q, and *New Round*, a binary variable that equals one if a firm receives a new round of VC financing in quarter q. In all models, the sample, independent variables, and coefficient interpretation are the same as the OLS models in Table VI. Independent variables are standardized for convenience and lagged one quarter. *LI Similarity* and *Foreign Similarity* are orthogonalized relative to *Private Similarity*. For brevity, control variables are omitted. 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)
Disruptive Potential	0.162***	0.670***	0.919***	0.466***	0.139***	0.162***
	(5.02)	(6.82)	(5.33)	(4.08)	(2.59)	(4.44)
Tech Breadth	-0.323***	0.017	0.309	-0.299	-0.164	-0.441***
	(-5.91)	(0.11)	(0.87)	(-1.59)	(-1.50)	(-6.65)
Private Similarity	0.060	1.247***	0.866**	1.265***	0.037	0.018
	(0.92)	(6.21)	(2.00)	(5.39)	(0.28)	(0.24)
LI Similarity	0.207***	0.477***	0.415	0.267	0.128	0.124*
	(3.52)	(2.76)	(1.15)	(1.27)	(1.12)	(1.70)
Foreign Similarity	-0.054	-0.025	0.380	-0.198	0.128	-0.088*
	(-1.30)	(-0.21)	(1.34)	(-1.42)	(1.38)	(-1.85)
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	342,146	347,918	114,364	233,550	112,643	229,499
Firms	9,145	9,167	3,972	7,543	3,951	7,483
R2 (%)	32.4	2.2	2.6	1.6	24.7	27.6

Table IA4: Subsample analysis of startups' exit: Time

This table repeats the OLS cross-sectional tests in columns (3)-(4) from Table VI on two subsamples. The tests relate a startup'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 VI. 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 startup's patents. Location fixed effects are based on the state reported in VentureXpert. Adjusted R^2 is reported as a percentage. Standard errors are clustered by startup 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)
Disruptive Potential	0.127***	0.089***	-0.186***	-0.025
	(2.73)	(4.41)	(-6.32)	(-0.98)
Tech Breadth	0.109	0.012	-0.286***	-0.179***
	(1.31)	(0.47)	(-3.91)	(-4.68)
Private Similarity	-0.017	-0.017	-0.164*	-0.466***
	(-0.16)	(-0.56)	(-1.76)	(-9.18)
LI Similarity	-0.001	-0.019	0.017	-0.062
	(-0.01)	(-0.64)	(0.24)	(-1.34)
Foreign Similarity	0.062	-0.008	-0.036	0.077***
	(0.87)	(-0.45)	(-0.66)	(2.69)
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
Observations	112,643	229,499	112,643	229,499
Firms	3,951	7,483	3,951	7,483
R2 (%)	0.3	0.5	0.4	0.6

Swiss Finance Institute

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

swiss:finance:institute

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