Internet Appendix for

"Rapidly Evolving Technologies and Startup Exits"

August 2021

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

IA.A Defining the Entity Type of Patents' Assignees

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

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

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

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

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

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

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

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

IA.B Matching patents to VentureXpert

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

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

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

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

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

[49] Firms can receive patents before VC funding.

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

[51] **?** 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

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.

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

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

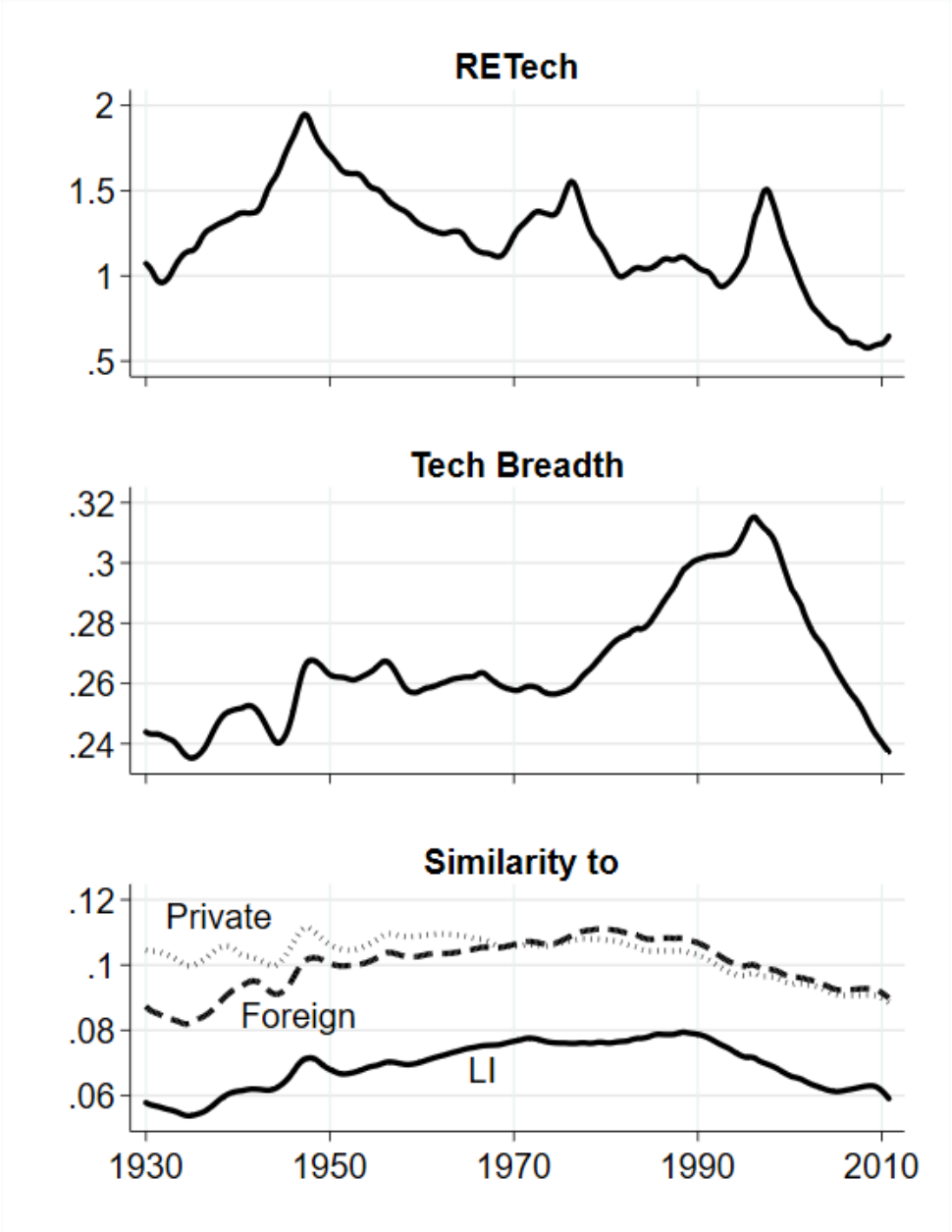


Figure IA.2: Counterfactual: Holding the Patent Wordspace Fixed

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

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

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

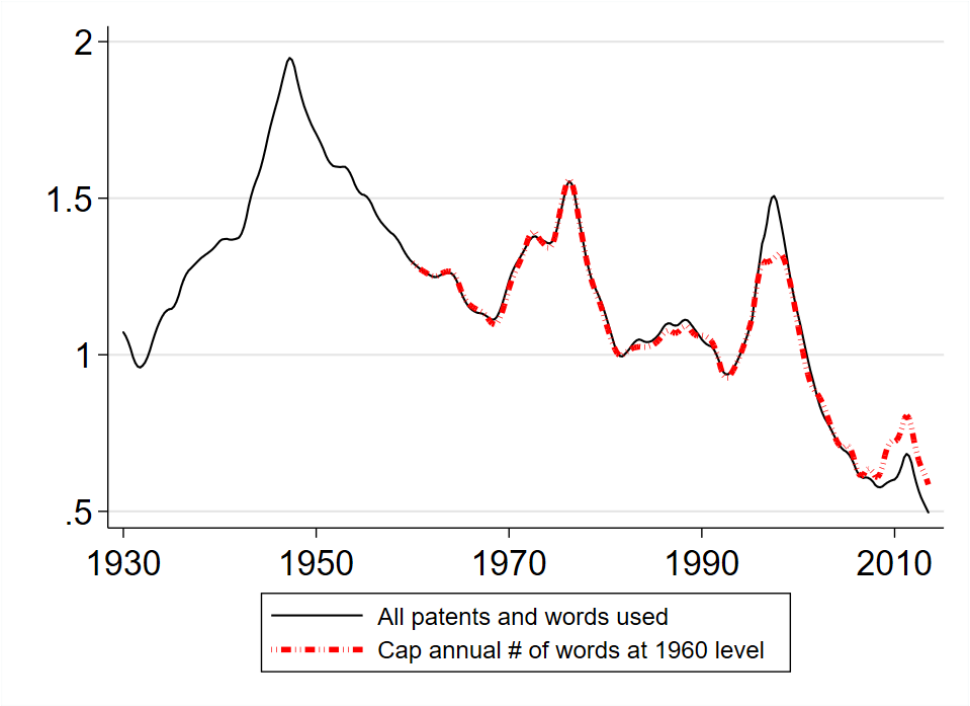


Figure IA.3: RETech After Stripping Fixed Effects

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

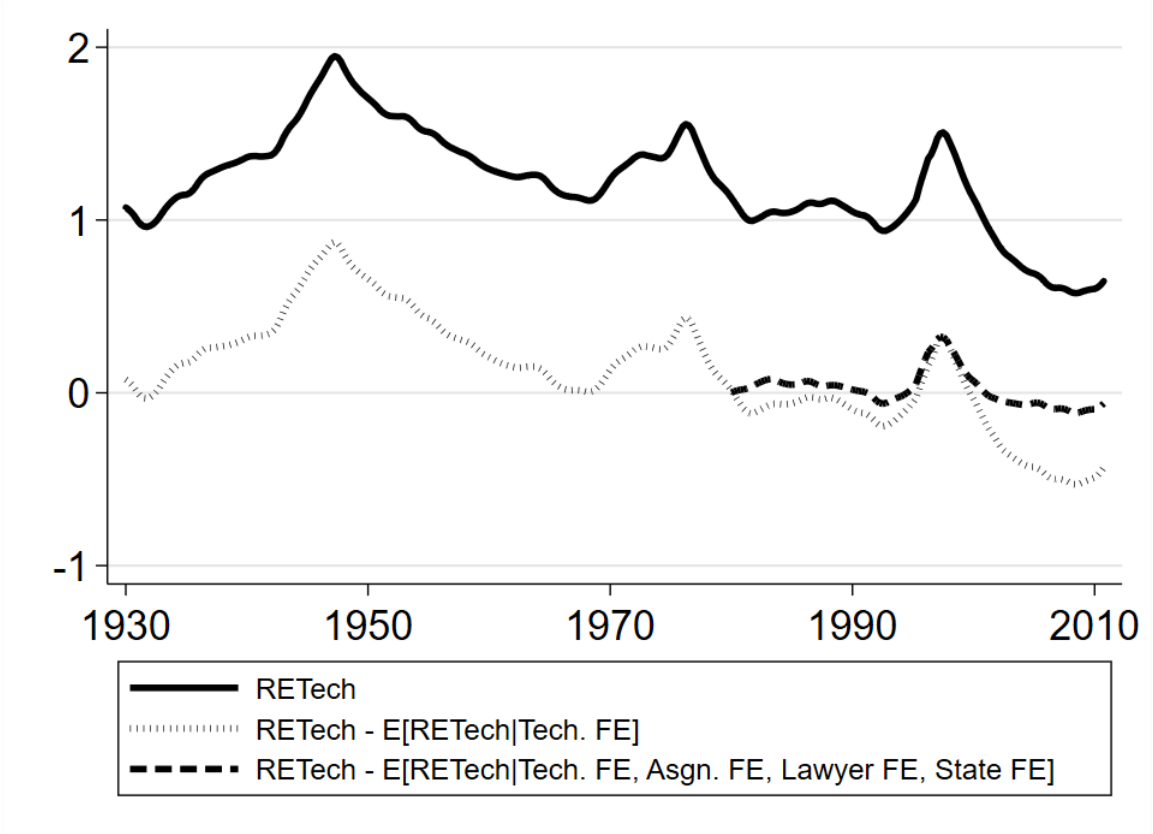
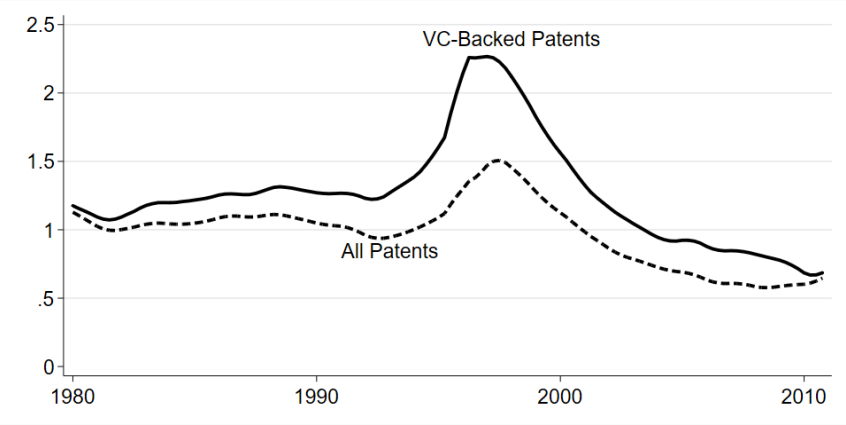


Figure IA.4: VC-Backed Startups: RETech and Exit Trends

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

Panel A: RETech – Startup Sample vs. Aggregate Data



Panel B: Exits – Startup Sample vs. Aggregate Data

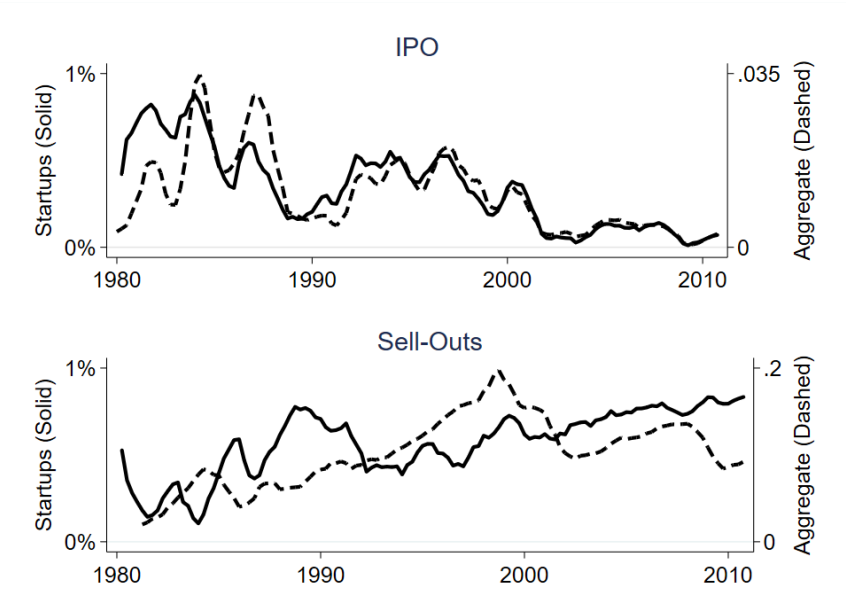


Table IA.1: Robustness of Baseline Results to Aggregration of RETech

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

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

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

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

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