Columbia Business School

COLUMBIA UNIVERSITY
Data Science Institute

COLUMBIA UNIVERSITY

Measuring Founding Strategy

Jorge Guzman (Columbia University)

Aishen Li (Columbia University & University of Michigan)

Strategy Science Conference

Philadelphia / Zoom

May 1st, 2020

What is strategy?

The organization of a firm in a way that predicts positive performance over the long run.

- It exists
- Managers can learn how to achieve it

SMS Strategic Management Society

STRATEGIC MANAGEMENT JOURNAL

SEJ STRATEGIC ENTREPRENEURSHIP JOURNAL

GLOBAL STRATEGY JOURNAL

ACADEMY OF Management

MANAGEMENT SCIENCE

STRATEGY SCIENCE

But how do we know if a company has a good strategy?

Notwithstanding tremendous interest in learning how to make strategy, and a clear definition, there is very little ability to score it.

This is even true within a specific school of thought

Positioning

Strategy is how to get to competitive advantage by occupying a lonely spot in the value proposition map.

“Competitive advantage is the fundamental basis [to achieve] above average performance in the long run". Porter (1985)

“Competitive strategy is about being different.” Porter (1996)

On the other hand, humans appear quite capable of scoring a strategy.

How to assess strategy?

They can listen, for example, to company statements, where companies emphasize their differences and unique value proposition (even within the same industry).

Consider Two Company Slogans

Southwest®

"Low fares. Nothing to hide. That's TransFarency!".

DELTA

"World's Most Trusted Airline"

Consider Two Company Slogans

Southwest®

"Low fares. Nothing to hide. That's TransFarency!".

Low cost and transparency. (cost leadership)

Appealing to cost-sensitive customers tired of extra fees.

DELTA

"World's Most Trusted Airline"

Trust and global coverage (variety and quality)

Appealing to those that want to get anywhere, reliably and on time.

Adding a third company

“Less Money, More Go”

A strategy analyst would quickly recognize:

1. Spirit Airlines also focuses on low-cost advantages.
2. Would compete more closely on the value proposition of Southwest, hurting its competitive advantage.
3. Southwest is less well positioned now than Delta (in this three firm comparison)

(this is not product differentiation, but value proposition!)

Now… expand on this idea

If an analyst was able to get the marketing materials for all airline carries, or even all U.S. companies.

- Could she systematically map the differences and ‘distance’ in the value propositions of firms?
- Wouldn’t a measure of how ‘far’ is a company from others reflect better (or worse) strategic positioning, even if imperfectly?

But, where to get marketing materials?

Key insight:

- All growth-oriented startups advertise their value proposition early on in the same channel: their website.

In this paper…

We develop a novel machine learning approach to measure strategy using the founding websites of all firms.

Paper in a nutshell

Approach

- Using all founding websites of startups in Crunchbase and contemporaneous public company 10K.
- We measure the distance in text of the words in these.
  - Most companies are obviously unrelated, but some do show important overlap.
- And then aggregate as a single *Founding Strategy Score* (how different it is from other firms at founding)

All software publicly posted here:

Results: founding strategy predicts performance

- firms with a higher founding strategy score are more likely to exit
- These differences are appreciable even within the seed financing round.

Founding strategy can be measured, and it partially determines firm performance.

- Introduction
- Theory
- Methodology
- Data
- Results
- Conclusion

Startups also position themselves at founding

"Who Are You?...I Really Wanna Know": Product Meaning and Competitive Positioning in the Nascent Synthesizer Industry

Callen Anthony[†]
Carroll School of Management, Boston College, Chestnut Hill, Massachusetts 02467, callen.anthony@bc.edu
Andrew J. Nelson
Lundquist College of Business, University of Oregon, Eugene, Oregon 97403, ajnelson@uoregon.edu
Mary Tripsas
Carroll School of Management, Boston College, Chestnut Hill, Massachusetts 02467, tripsas@bc.edu

- Introduction
- Theory
- **Methodology**
- Data
- Results
- Conclusion

Setup

- A startup is born in a world with incumbents, indexed by $j \in \{1, \dots, J\}$
- The value proposition of the startup has an elasticity of substitution with each incumbent ϵ_j
- Leading to market power

$$M = g(\epsilon_1, \dots, \epsilon_J)$$

- The firm realizes profit based on market power, Demand D, and some random error

$$\pi = h(M, D)\mu$$

Definition: The Strategy Score

Definition. ***Strategy Score:***

- *For any company, a measure of how good their strategy is (their 'strategy score') $S > 0$ is a scalar measure that can be positive and monotonically translated to higher market power through some positively increasing function f*

$$M = \zeta(S), \qquad \frac{\partial \zeta}{\partial S} > 0 \; \forall S$$

- ***The goal of our paper is to develop an empirical approach to measure S.***

Three insights to measuring founding strategy

Insight 1.
Startups state their value proposition

While it is virtually impossible to measure the value the consumer sees on a firm (and the substitutability across firms, i.e. the ϵ_j), it is possible to see what the firm *thinks* its value is.

Assuming a certain level of similarity between the two.

Similarity

- For two statements s_i and s_j by a startup and a incumbent

$$\sigma_{ij} = h(s_i, s_j),\ \sigma_{ij} \in [0,1]$$

- Distance

$$d_{ij} = 1 - \sigma_{ij}$$

Insight 2.
one type of founding statement can be found for most startups

Their website.

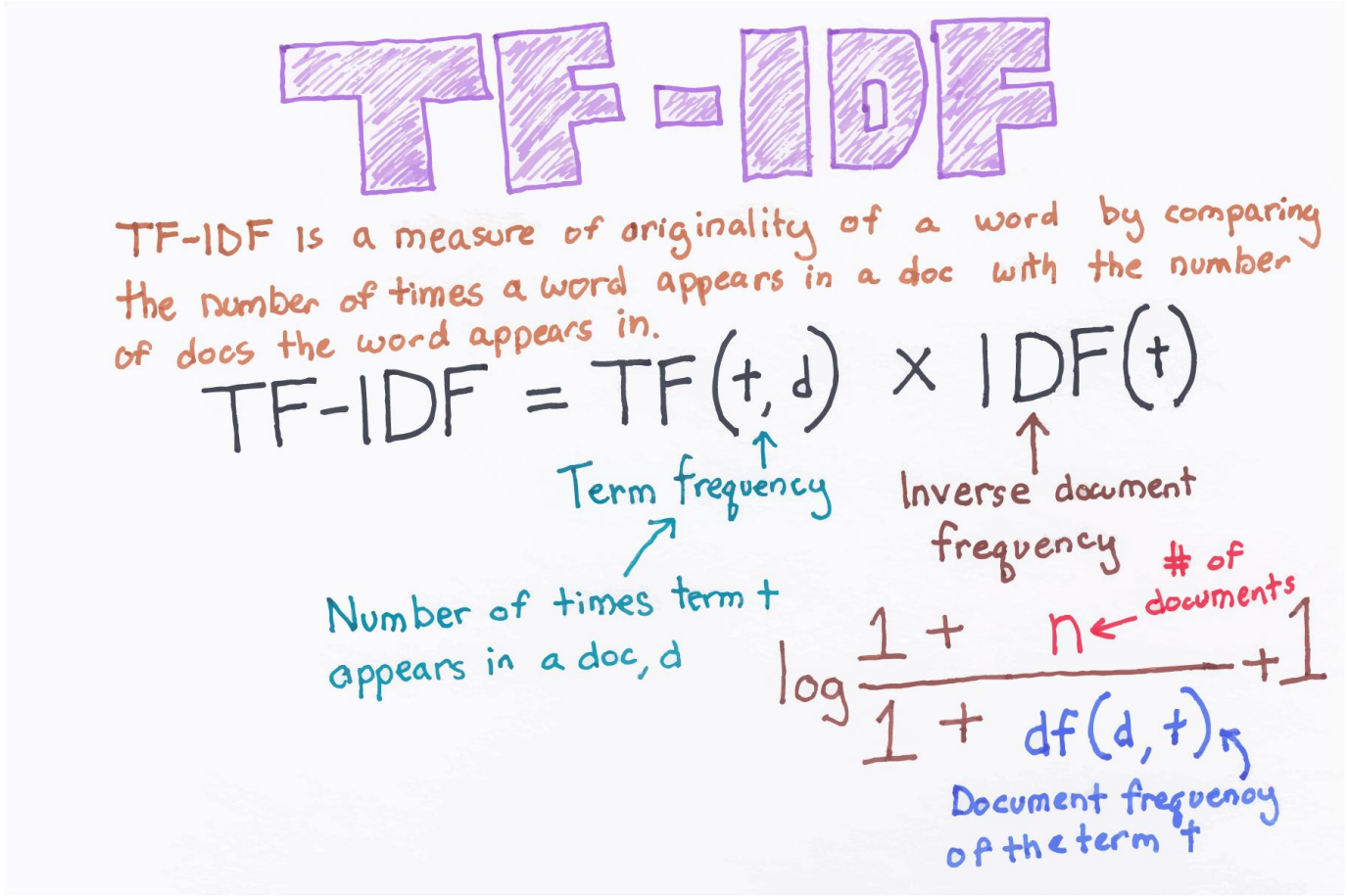
Startups trying to help consumers and other audiences learn about them state carefully what they are doing, to the best of their abilities.

(much more could be done to learn about audiences!)

e.g. Bourveau & Breuer

Insight 3.

Similarity in text can be measured through machine learning



From similarity to a strategy score

How to aggregate a pair-wise measure into a single score of positioning?

- Strategy theory emphasizes how ‘lonely’ you are, focusing specifically in the few firms that are close.

- In industrial organization, we find oligopolistic rents decrease quickly and mostly disappear after 4 or 5 competitors.

Entry and Competition in Concentrated Markets

Timothy F. Bresnahan and Peter C. Reiss
Stanford University

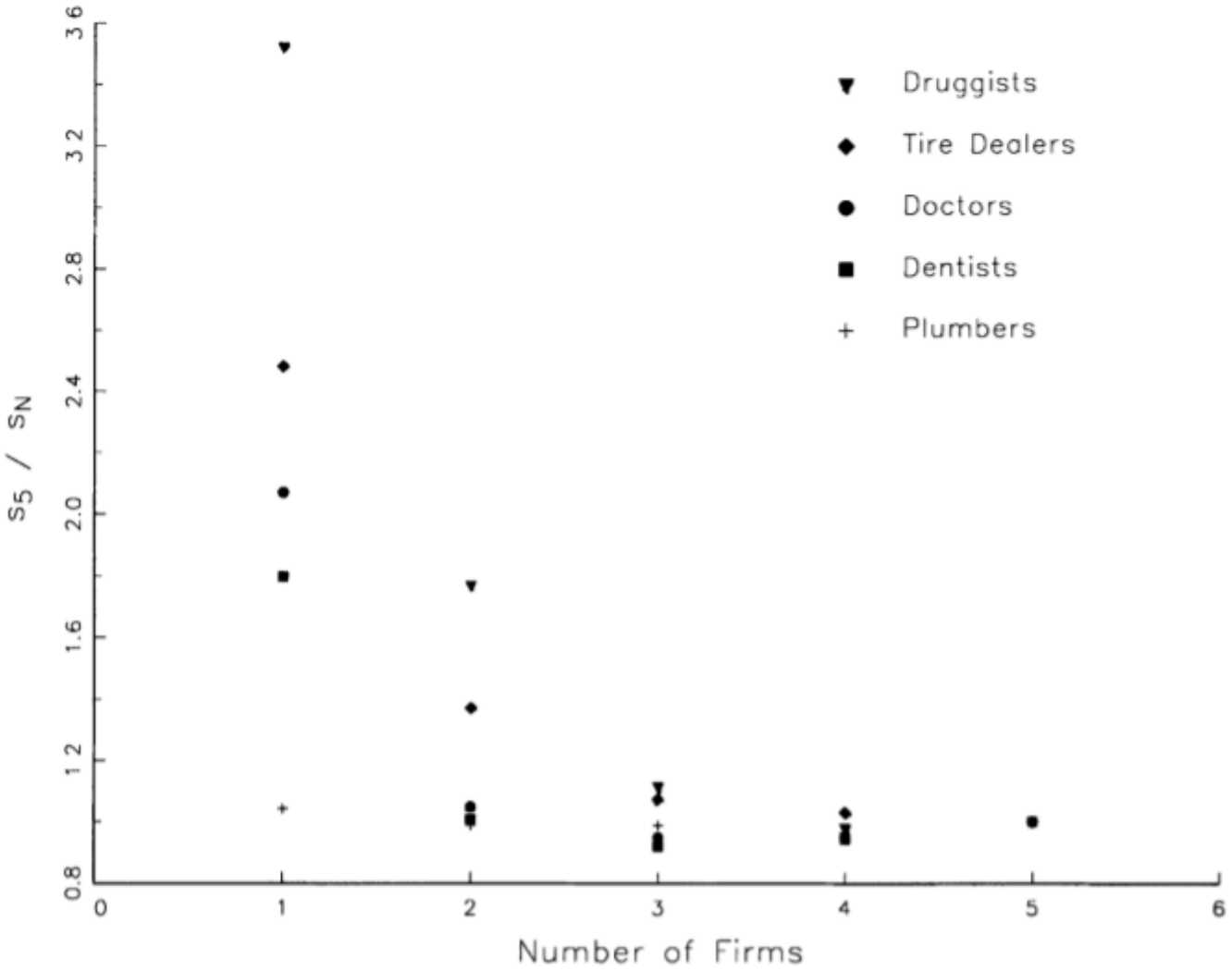
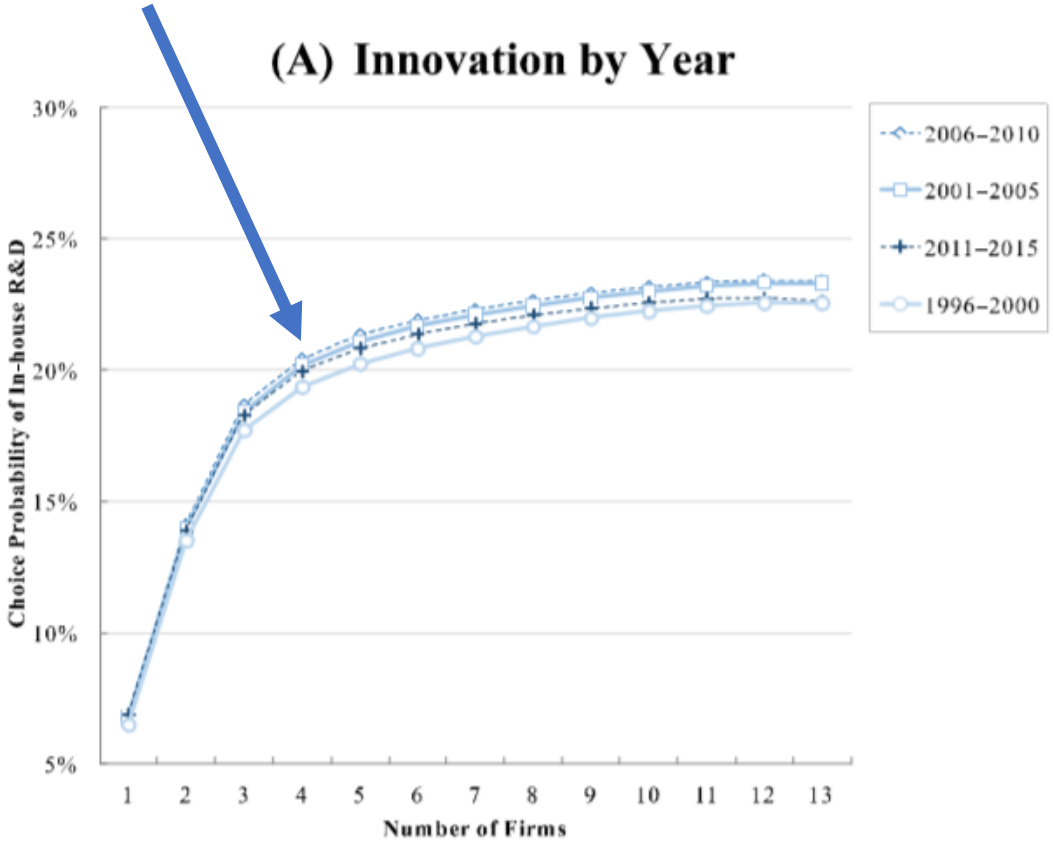
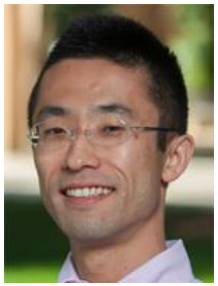


FIG. 4.—Industry ratios of s_5 to s_N by N

Also on role of competitive pressure to innovate



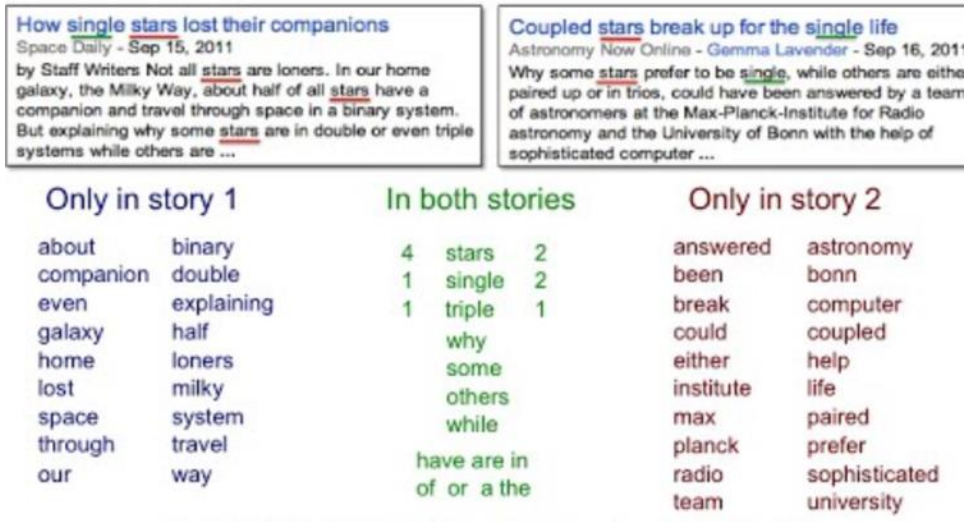


Mitsuri Igami

Kosuke Uetake

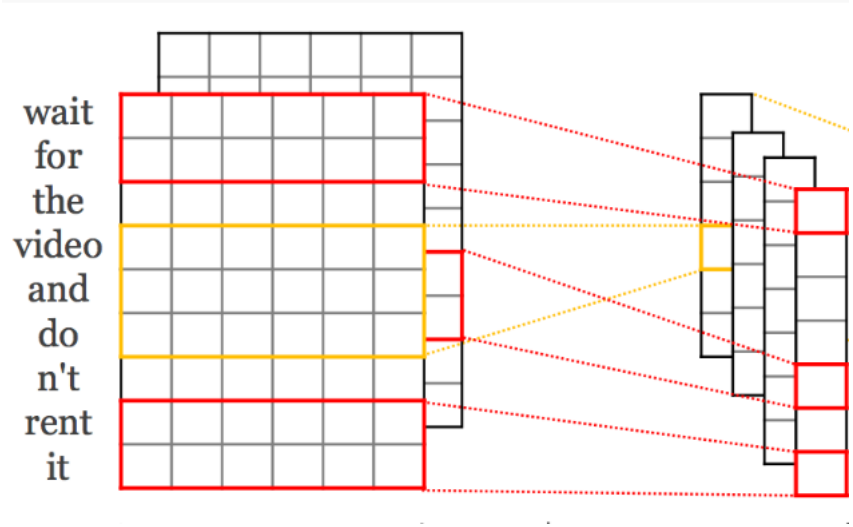
Quantifying text: text-as-data

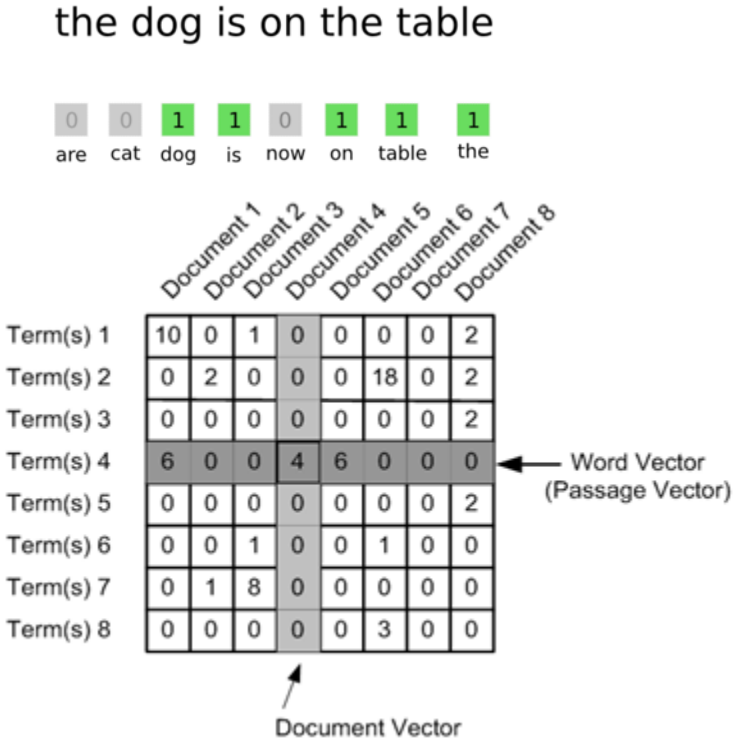
1. **Create a *clean* corpus**
   - No stop words
   - Only nouns
   - Appear in less than 20% of total documents



2. Embed the document as a vector

- For each document, do a frequency count for each words in corpus
- Represent each document in a mathematical way

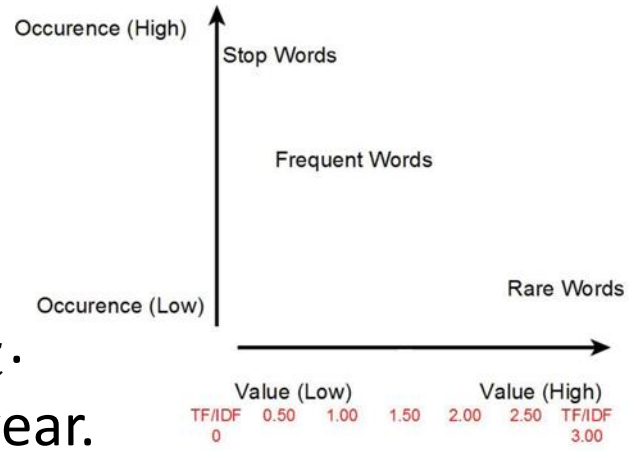




3. Weight using TF-IDF algorithm

- Change the entries of document vector to weights instead of word frequency.
- Give higher weights for rarer words.
- Term frequency count is compared to an inverse document frequency count, which measures the number of occurrences of a word in the entire corpus.

$$W_{ij} = tf_{ij} \times \log_2 \frac{N}{n}$$



W_{ij} = Weight of term in document D_i .

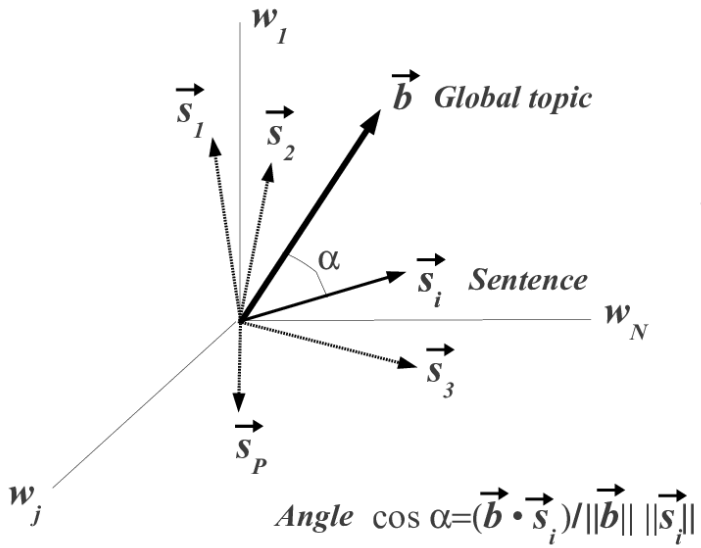
tf_{ij} = Frequency of term T_j in document D_i.

N = Number of documents in the specific year.

n = Number of documents that contains T_j.

Compare similarity

- Document vectors are in very high dimension spaces.
- The vectors are sparse (0 in most of entries), they carry the information about a companies product.



- Comparing the angle between each two vectors can offer us their similarity. (Cosine similarity)
- 1 is two equal statements, 0 is two completely different statements.

Text-Based Network Industries and Endogenous Product Differentiation

Journal of Political Economy, Forthcoming

69 Pages
Posted: 9 Dec 2009
Last revised: 4 Jul 2015

Gerard Hoberg
University of Southern California - Marshall School of Business - Finance and Business Economics Department

Gordon M. Phillips
Dartmouth College - Tuck School of Business; National Bureau of Economic Research (NBER)

Date Written: July 3, 2015

Key differences to Hoberg and Phillips

- Introduction
- Theory
- Methodology
- **Data**
- Results
- Conclusion

We bring together three distinct datasets

Growth Oriented Startup companies:

Crunch Base

Startup Historical Statements:

Way back Machine

Incumbent Performance:

Contemporaneous 10-K statements

While prior research has substantively used each of these datasets, often together, we provide the first automated approach to do so at scale

Growth Oriented Startup companies: crunchbase.com

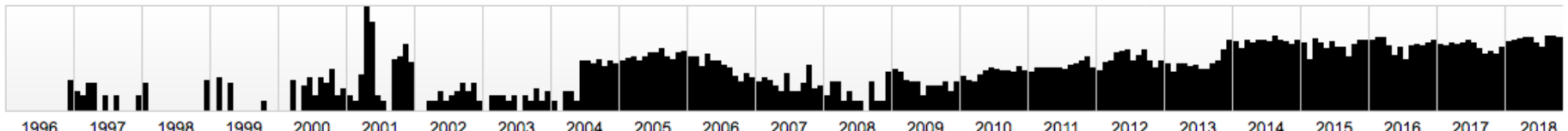
- A list of *most* venture growth oriented startups in the US.
  - Includes founding year, funding progress, and outcomes.
- We subset of U.S. companies that have the website reported in the data, address and financing records.

Startup statements at founding: Their website

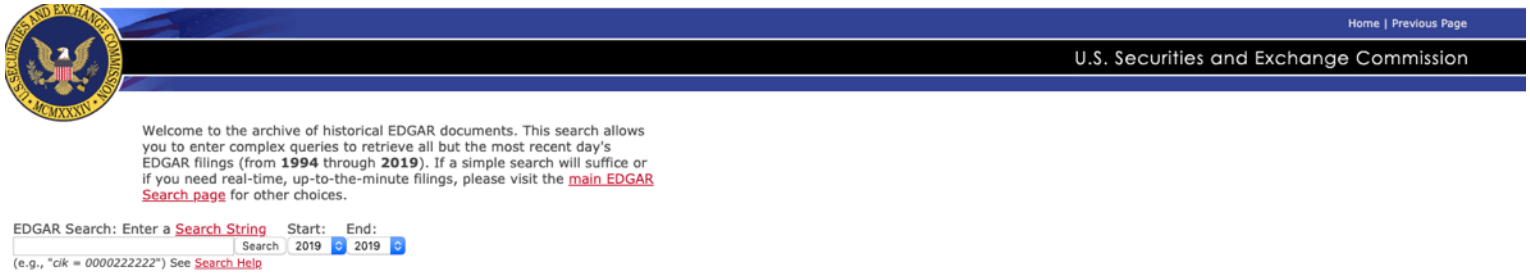
- The Internet Archive keeps digital historical copies of all websites in the internet.
  - All HTML tags and free text
  - All images
  - All publicly accessible unique links.

- It is free to scrape.

- Developed a scraper that finds the first (after founding) About Us or Product page in their historical website.

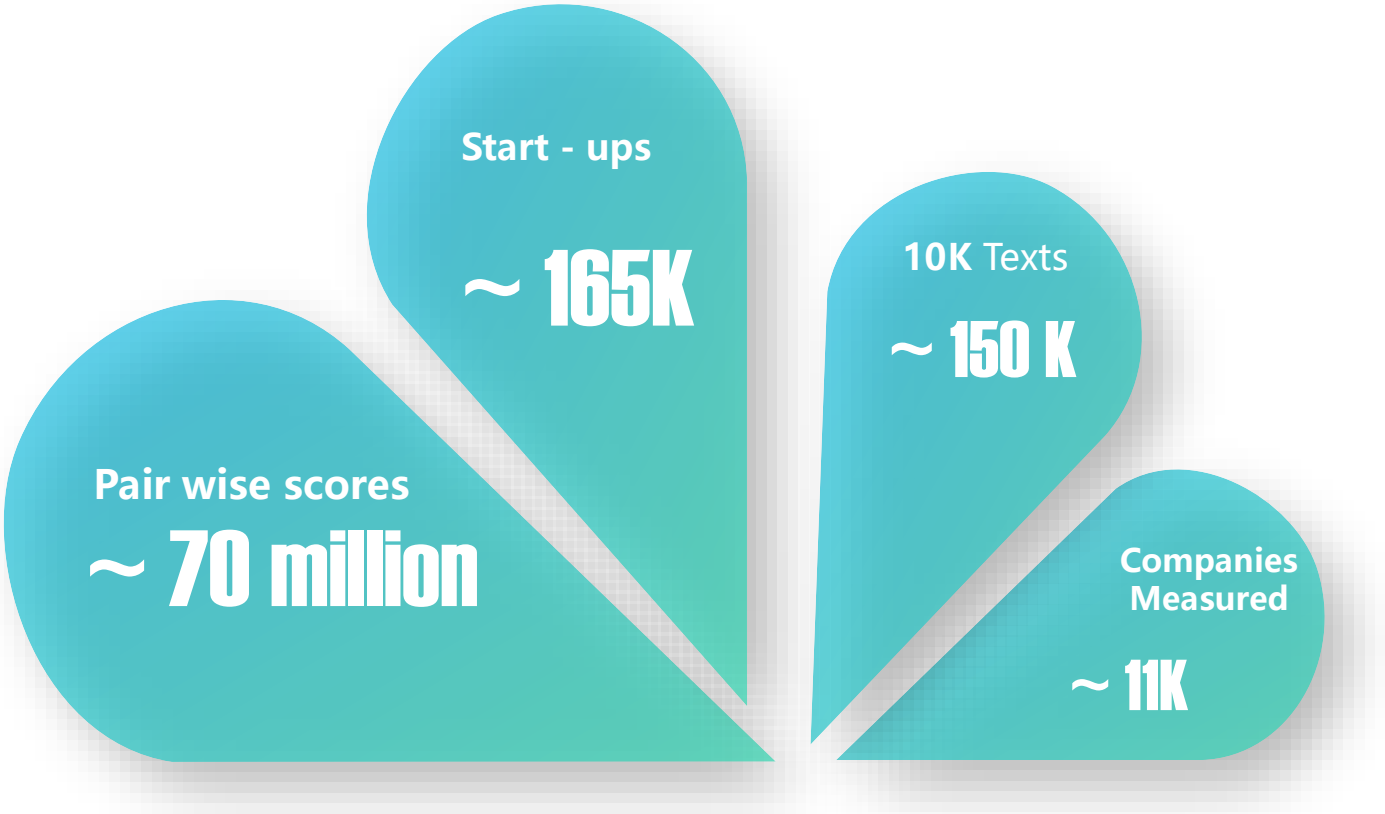


Public incumbent data: 10-K Statements



- All public companies are required to include in their annual report a “business description” section.
- SEC achieve has all required annual reports since 1997
- We use the business self-proposal part of text from each “annual report to shareholders” (Form 10-K)

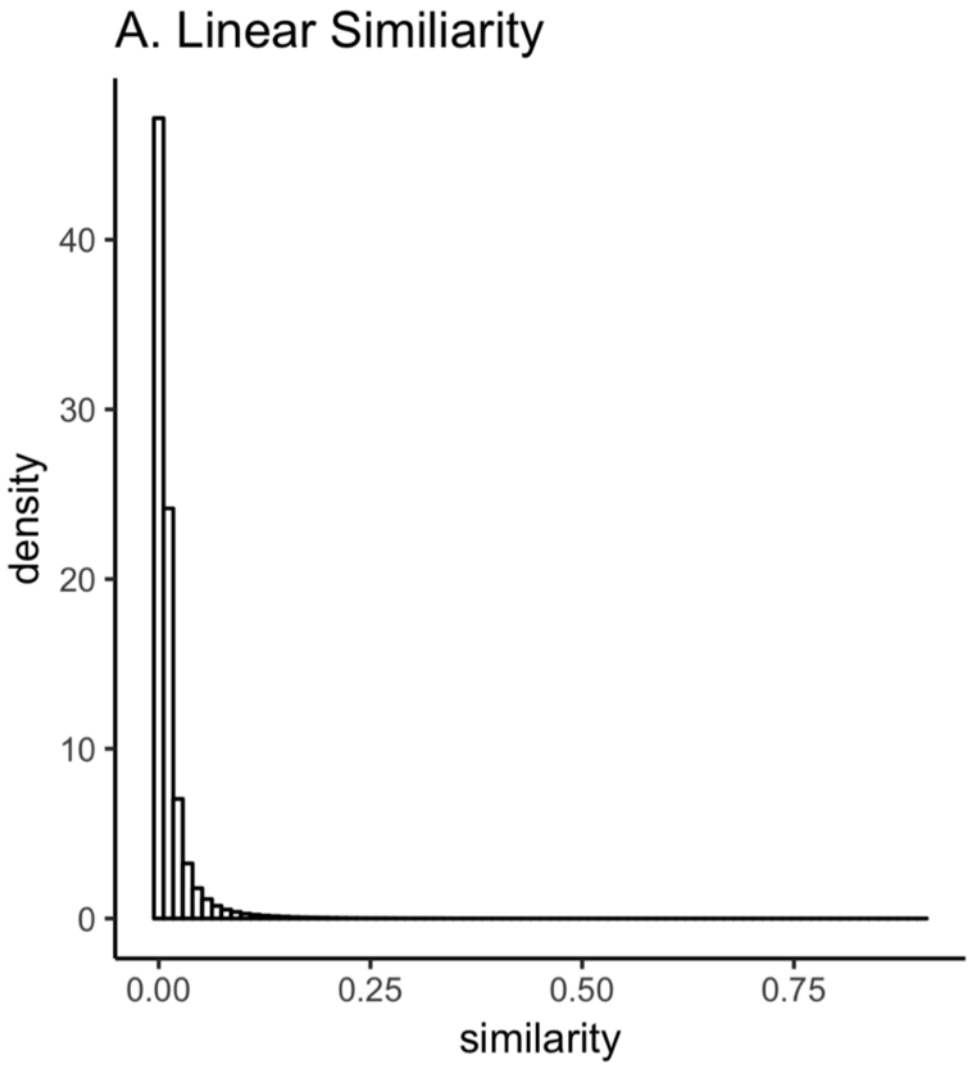
Text data-set summary



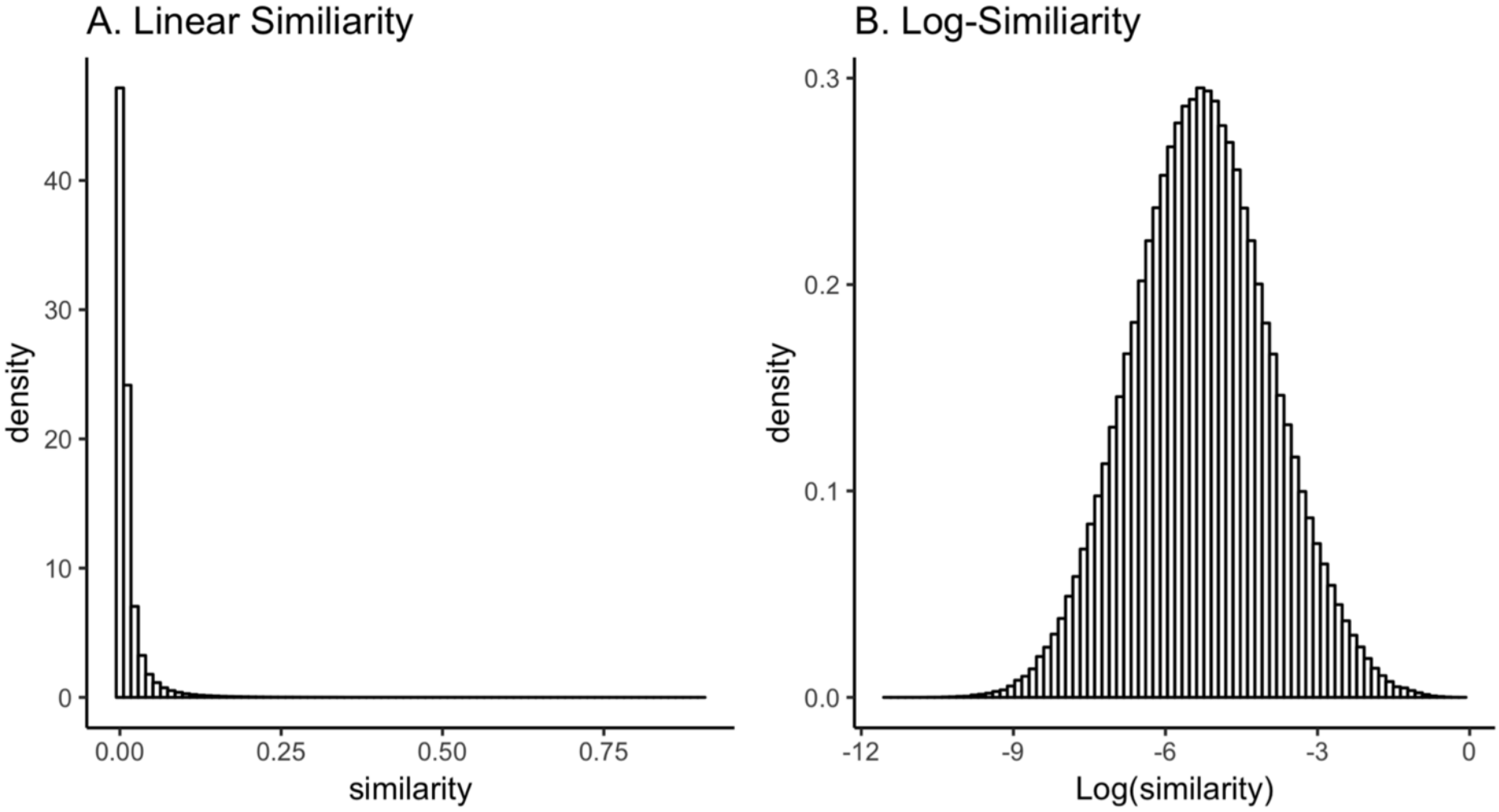
Contents

- Introduction
- Theoretical overview: Why measure startup strategy?
- Measuring startup strategy: a text-based approach
- Data
- **Results**
- Conclusions

Distribution of Similarity Scores Between Startup Websites at Founding and Contemporaneous 10-K Statements



Distribution of Similarity Scores Between Startup Websites at Founding and Contemporaneous 10-K Statements



Panel A: Summary statistics for similarity scores

Statistic	Mean	St. Dev.	Median
Similarity (67 million pairs)	0.0128	0.0263	0.0050

Summary Statistics

Panel A: Summary statistics for similarity scores

Statistic	Mean	St. Dev.	Median
Similarity (67 million pairs)	0.0128	0.0263	0.0050

- Now put these similarity measures together into a *Founding Strategy Score*

$$\widehat{S_i} = \frac{1}{5} \sum_{\{j \in J^5\}} (1 - \sigma_j),$$

$$J^5 = \{5\ closest\ incumbents\}$$

Founding Strategy Scores

Table 2: Estimated Score Summary Statistics

Panel A: Similarity Score Summary Satistics

Statistic	Mean	St. Dev.	Median
Similarity (67 million pairs)	0.0129	0.0264	0.0050

Panel B: Founding Strategy Score Summary Statistics

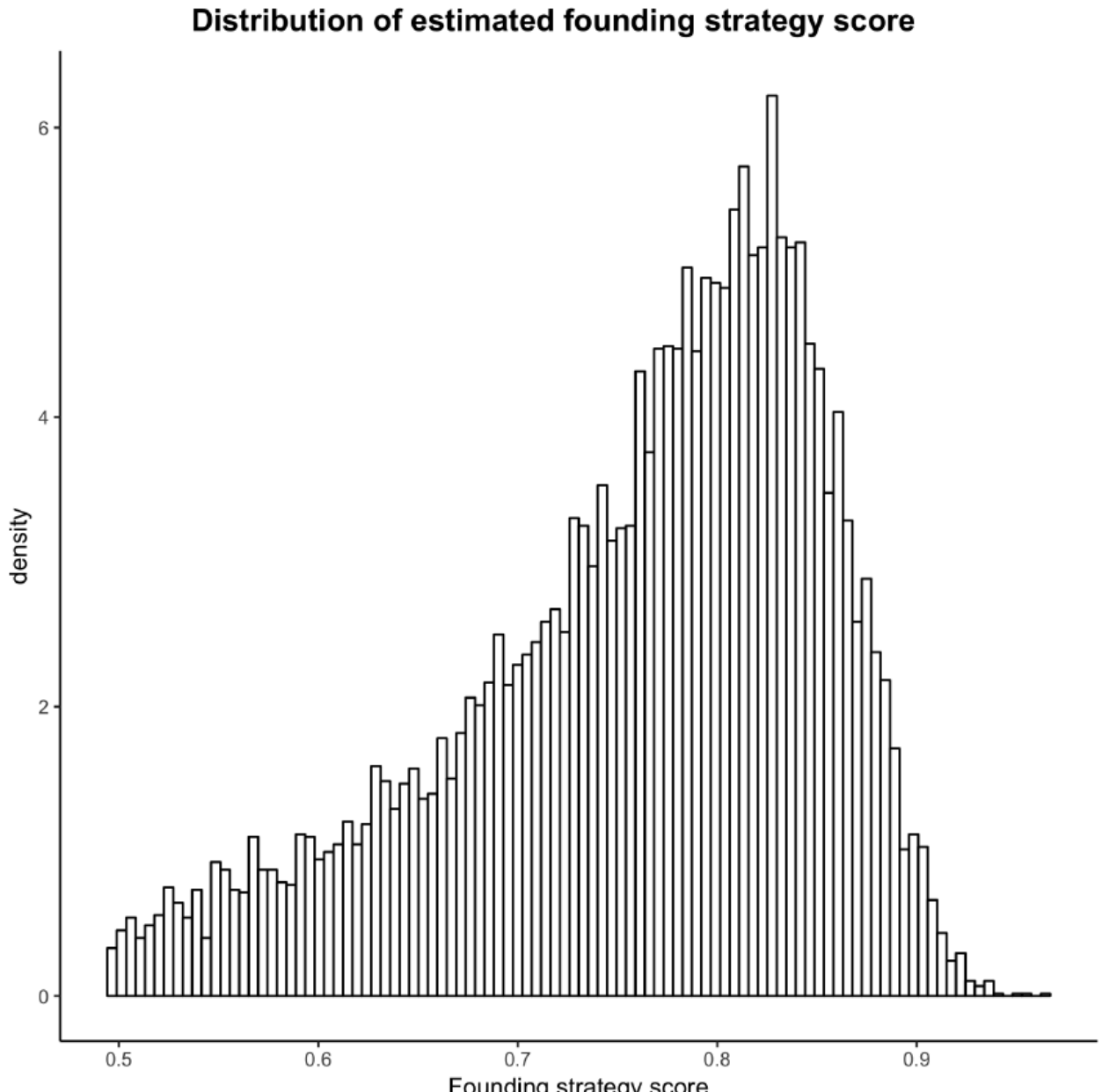
Statistic	Mean	St. Dev.	N
Founding Strategy Score (7 closest)	0.78	0.09	12,103
Founding Strategy Score (6 closest)	0.77	0.09	12,103
Founding Strategy Score (5 closest)	0.76	0.09	12,103
Founding Strategy Score (4 closest)	0.75	0.10	12,103
Founding Strategy Score (3 closest)	0.74	0.10	12,103
Weighted Strategy Score (5 closest)	0.76	0.10	12,103

How many close competitors?

Table 3: Correlation of Founding Strategy Scores

	(1)	(2)	(3)	(4)	(5)	(6)
(1) Founding strategy score (7 closest)	1					
(2) Founding strategy score (6 closest)	0.996	1				
(3) Founding strategy score (5 closest)∗	0.990	0.995	1			
(4) Founding strategy score (4 closest)	0.976	0.988	0.994	1		
(5) Founding strategy score (3 closest)	0.950	0.967	0.979	0.994	1	
(6) Weighed by market value (5 closest)	0.909	0.915	0.919	0.917	0.907	1

The Shape of Founding Strategy Scores



Founding Strategy Score Examples

(Less differentiated) *(More differentiated)*

Student loans
0.64

0.75

0.86

0.84

Founding Strategy Score Examples

Top Scores

0.96

0.93

Bottom Scores

0.60

0.60

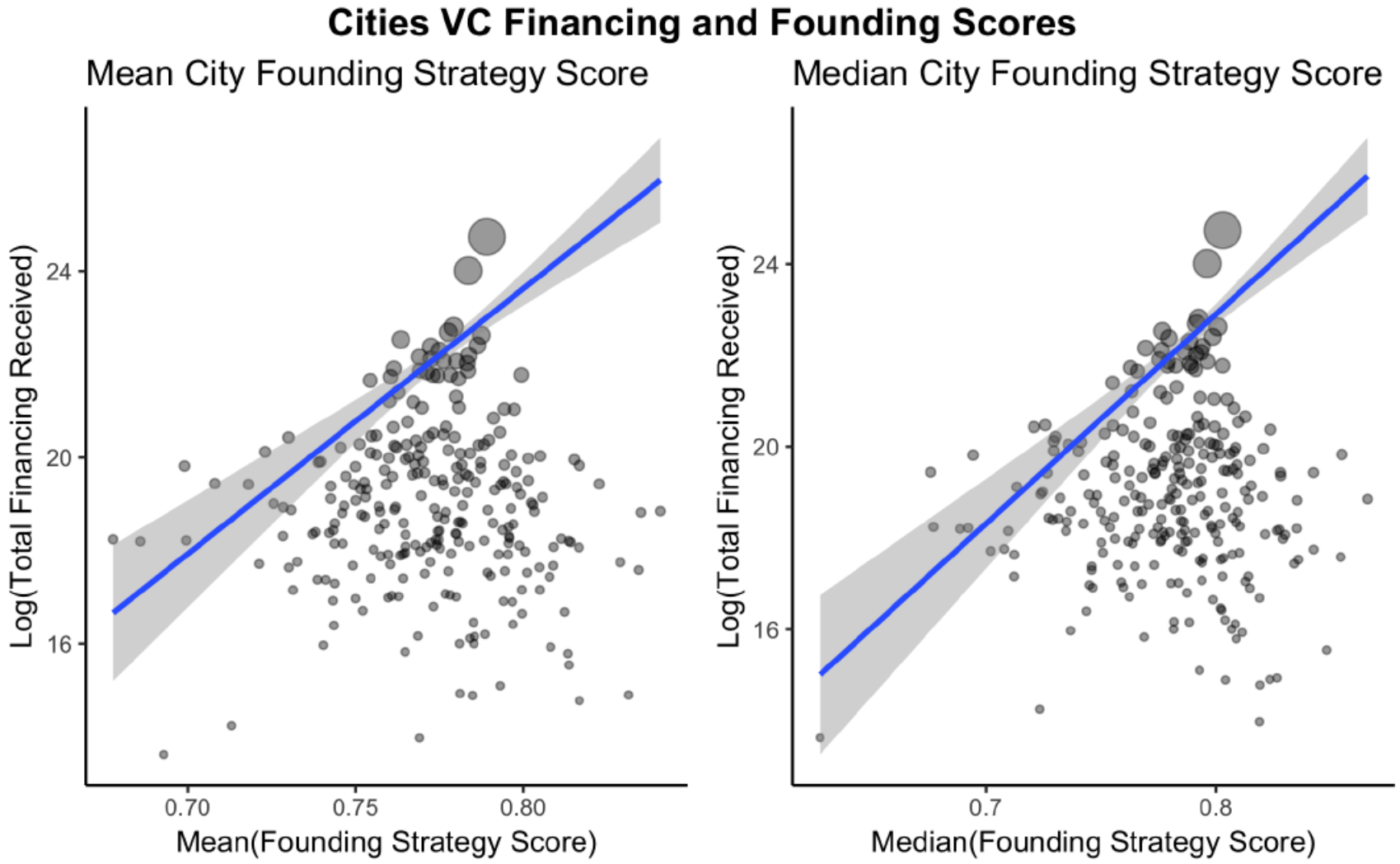

Cities VC Financing and Founding Scores
Mean City Founding Strategy Score
Median City Founding Strategy Score
Log(Total Financing Received)
24
20
16
0.70
0.75
0.80
Mean(Founding Strategy Score)
0.7
0.8
Median(Founding Strategy Score)

Table 4: Founding Strategy Score and Startup Performance

	Dependent variable:				
	Equity Growth (IPO or Acquisition)				
	OLS	OLS	OLS	OLS	Logit
	(1)	(2)	(3)	(4)	(5)
Founding strategy score	−0.073**	0.074**	0.048*	0.066**	0.580**
	(0.037)	(0.033)	(0.028)	(0.030)	(0.282)
Founding Year F.E.	No	Yes	Yes	Yes	Yes
City F.E.	No	No	Yes	Yes	No
Industry F.E.	No	No	No	Yes	No
Observations	12,103	12,103	12,103	12,103	12,103
R^2	0.0003	0.097	0.191	0.229	
Log Likelihood					−4,919.094

*Note: *p<0.1; **p<0.05; ***p<0.01*

Figure 4: Smoothed Relationship of Founding Strategy Score and Firm Performance

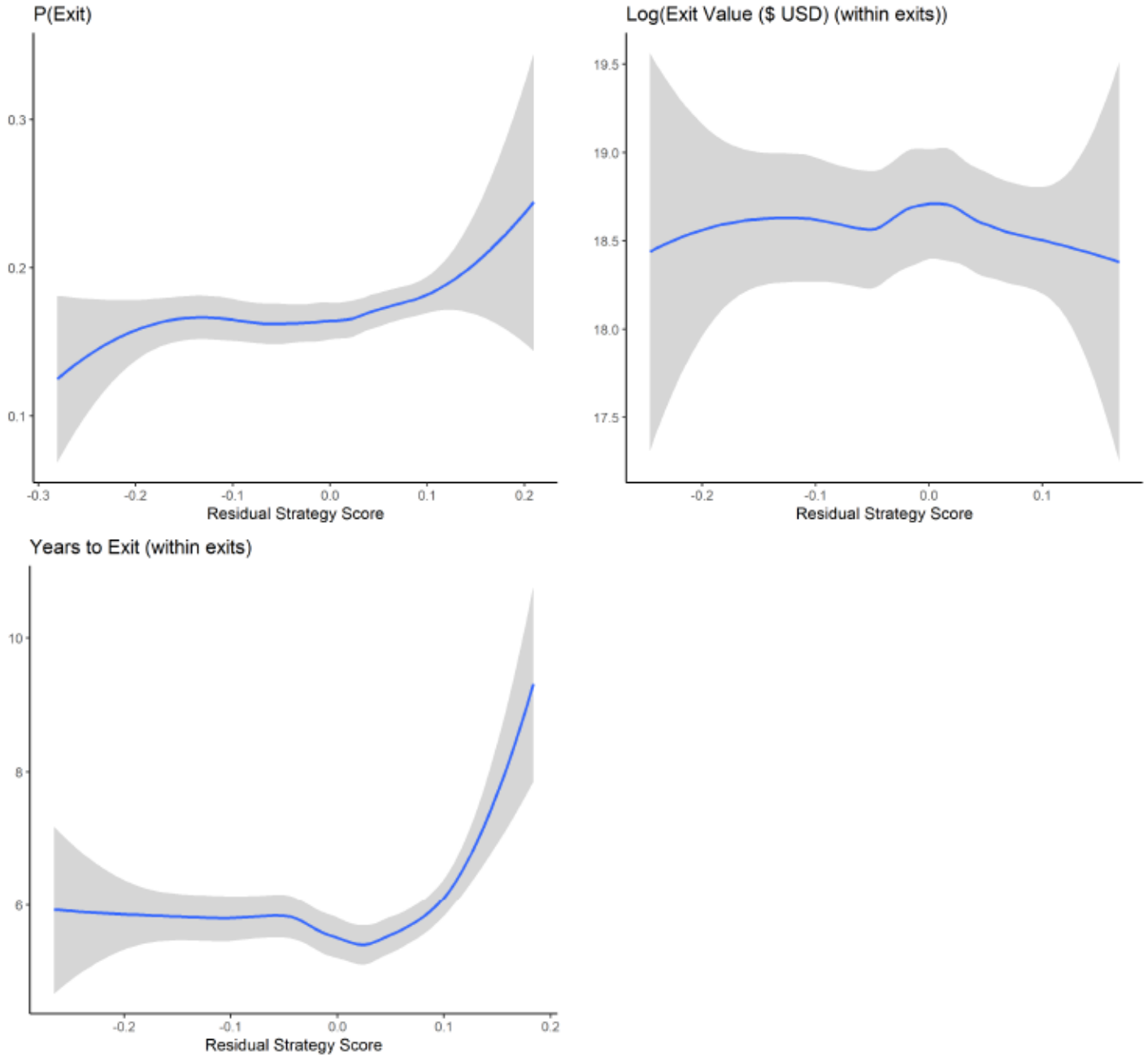


Table 5: Founding Strategy Score and Initial Financing

	Dependent variable:				
	Log(Seed Funding)				
	(1)	(2)	(3)	(4)	(5)
Founding Strategy Score	0.567**	0.370	0.367**	0.512***	0.289*
	(0.229)	(0.235)	(0.169)	(0.143)	(0.149)
Sample	All	All	All	First Event	All
Founding Year F.E.	No	Yes	Yes	Yes	Yes
City F.E.	No	Yes	Yes	Yes	Yes
Seed Funding Year F.E.	No	No	Yes	Yes	Yes
Industry F.E.	No	No	No	No	Yes
Observations	9,117	9,117	9,117	6,651	9,117
R^2	0.001	0.182	0.223	0.251	0.283

*Note: *p<0.1; **p<0.05; ***p<0.01*

Contents

- Introduction
- Theoretical overview: Why measure startup strategy?
- Measuring startup strategy: a text-based approach
- Data
- **Results**
- Conclusions

Conclusions

- Better strategies are those that make firms more differentiated, thus allowing it to charge higher profits.
- We develop a novel approach to measure founding strategy
  - Proposed an estimate – the founding strategy score
  - And an approach to do it using public statements by startups and incumbents
- Our measure predicts three empirical regularities consistent with a measure able to score founding strategy.

Thank you!!