Measuring Founding Strategy*

Jorge Guzman
Columbia University †

Aishen Li
Columbia University ‡

Abstract

We introduce a novel approach to measure startup founding strategic differentiation and study how it predicts follow-on performance. We use natural language processing and historical websites to estimate the similarity between the founding website of an individual startup, the historical website of public firms at the startup's founding year, and the founding website of other startups founded in the same year. We propose that distance in the value proposition stated in these websites represents differentiation in the market. Startup differentiation is estimated as the average text-based distance from the five closest incumbents (public firms). We implement this approach using a large sample of startups in Crunchbase. Consistent with a measure of founding differentiation, we demonstrate that our measure predicts a meaningful increase in early stage financing and equity outcomes, both unconditionally and controlling for cohort and industry. The positive benefits of equity outcomes only evidence themselves after year 5 of age, suggesting more differentiated firms may take longer to prove themselves. Using out of sample tests, we also demonstrate that our measure is economically important, predicting 25% of total variation in the receipt of early stage financing and 21% of variation in equity outcomes. This, in turn, implies that founding strategy is a key element of startup performance. Open source code to download websites and to estimate founding differentiation is provided.

Keywords: *Entrepreneurship , Strategy , Strategic Positioning , Machine Learning , Text Analysis*

*We thank Mabel Abraham, Eric Abrahamson, Natarajan Balasubramanian, Joel Brockner, Raj Choudhury, Bo Cowgill, Enrico Forti, Kathy Harrigan, Daniel Keum, Bruce Kogut, Kristina McElheran, Stephan Meier, Patryk Perkowski, Robert Seamans, Scott Stern, and Zhiyuan Yu for helpful comments, as well as participants in the Columbia Data Science Day, the NYU Workshop on Machine Learning and Strategy, and Strategy Science.

†Jorge Guzman: Uris Hall, 711. Columbia University. Broadway and 116th. New York, NY. 10027. jag2367@columbia.edu.

‡Aishen Li: Is a graduate from Columbia University. He will be starting his PhD at University of Michigan in the summer of 2021. al3836@columbia.edu.

1 Introduction

According to research tradition, the defining characteristic of the entrepreneurial environment is the presence of change (Schumpeter 1942). In the classic works of Nelson & Winter (2002), Baumol (1968), and Aghion & Howitt (1992), the role of the entrepreneur is to face the changing market structure of an economy and advance from one moment of economic equilibrium to the next. Knight (1921) similarly emphasized that the distinctive characteristic of the entrepreneur is to deal with undiversifiable risk. In general, uncertainty and change are consistently characterized as the essence of entrepreneurial activities (Shane 2003, Lerner & Schoar 2010).

Recognizing the importance of uncertainty has led to a view that successful entrepreneurial strategies should emphasize adaptability—being able to deal with change and having a process to advance in it. Eisenhardt & Martin (2000) and Teece et al. (1997) call organizational routines that allow doing so 'dynamic capabilities', and posit that they constitute the central element of success in fast changing industries (see also Winter 2003, McGrath & MacMillan 2000). Recently, research has coalesced around one specific dynamic capability as key for entrepreneurship: experimentation, or the ability to validate hypotheses by gathering information and then proceeding to undertake strategic commitments (Koning et al. 2019, Gambardella et al. 2018, Reis 2011). While dynamic capabilities matter in many types of firms, Teece et al. (1997) argue they are are most important in technology entrepreneurship.

In contrast to this perspective, other work has emphasized that there are static characteristics of a startup's strategy, often present at founding, that play a key role in determining follow-on performance. In a study of venture-backed firms, Kaplan et al. (2009) show that there tends to be little change between the value proposition included in startup's business plans when initially approaching venture capital investment and in their eventual IPO prospectus. Guzman & Stern (2019) show that the characteristics that are observable at the very moment of a firm's initial legal registration substantially predict follow-on performance. Barney (1991) emphasizes that it is underlying stable firm resources, such as intellectual property, that are central to sustained competi-

tive advantage (see also Hellmann & Puri 2000). And multiple ecological perspectives argue that founding conditions partially determine firm evolution (Stinchcombe 1965, Hannan & Freeman 1977, Moreira 2016). Consistent with these results, Porter (1996) considers 'market positioning', the ability to occupy differentiated locations in a value proposition map, the essence of competitive strategy. Qualitative work validates its importance. In an in-depth analysis of the synthesizer industry, Anthony et al. (2016) show that the initial messaging of startups on their value proposition is key to their eventual evolution and market performance.

Of course, dynamic capabilities and founding positioning are not exclusive to each other. Both founding strategies and dynamic capabilities are likely to matter, and may even support one another, in startups. The recent entrepreneurial strategy framework by Gans et al. (2019), for example, recognizes the importance of both, and develops a cohesive set of tools to use founding positioning and experimentation together. A more appropriate and relevant question is how founding differentiation relates to performance and its relative value—how and how much does it matter. However, there appears to be little ability so far to investigate these questions within strategy because of the inherent difficulties in measuring the strategic positioning of firms in general, even more so for very young startups. How may one statistically characterize the difference in the level of founding differentiation for startups? How would we measure its role in predicting startup performance outcomes? More generally, how do you measure founding strategy?

The purpose of this paper is to introduce a systematic approach to measure the founding differentiation of startups and assess its role in predicting performance. To do so, we take advantage of natural language processing methods that incorporate not only the incidence of words, but the context in which they are used (word embeddings), to characterize the distance in market positioning across all startups and incumbents at the time of startup founding. Using the WaybackMachine, an online historical archive of internet web pages, we download the website at or close to founding for a large sample of startups from Crunchbase, and the website of all public companies during the startup's founding year. We then estimate the distance of the stated value proposition in each website to all others, and aggregate to only consider those incumbents and

other startups that are closest to the focal firm. This approach allows us to create a new measure, *Differentiation Score*. The differentiation score represents the conceptual distance between the value proposition stated by a startup in its website at founding, and the value proposition stated by other companies in the market. Building from findings in industrial organization that the five closest competitors shape profits (Bresnahan & Reiss 1991, Igami & Uetake 2019), we focus on the average distance of the focal startup to the five closest incumbent public firms at founding.

After developing our measure of founding differentiation, we proceed to study its role in predicting performance. We emphasize that these estimates are not causal, but instead comparative statics of the relationship between founding strategy and eventual firm success. They are validations of our measure and descriptive correlations of the way founding strategy predicts success. We focus on three key results.

First, we consider how founding strategy predicts early stage financing. After including founding year and industry fixed effects,[1] we document a strong relationship of our measure to this outcome. Moving from the 10th to the 90th percentile of our measure predicts 25% higher early stage financing (defined as the sum of seed capital, grants, and angel financing) and 49% higher series A financing. In contrast—and consistent with the idea that we are capturing positioning vis-á-vis the consumer market and not the financing market—differentiation from other startups at founding does not relate positively to performance, and is negative under some specifications.

Second, we consider how founding strategy predicts long-term equity outcomes such as IPO or acquisition. We report that our differentiation score also predicts a higher probability of IPO or acquisition, though the relationship here is more nuanced. There is only a weak positive relationship of differentiation to equity outcomes. Yet, this relationship strengthens when dropping very young firms for which we do not observe later years and firms with low value acquisitions. These patterns suggest a strong age dependency. Consistent with the existing literature documenting that more unique startups will struggle to achieve legitimacy initially, but may ultimately perform better (Deephouse 1999, Marx et al. 2014), there is a positive relationship between founding dif-

[1] We control for industry by using the time-varying text-based industry measure of Hoberg & Phillips (2016).

ferentiation and equity outcomes, but it does not evidence itself until year 5 of age. Afterwards, however, it is substantial. Moving from the 10th to the 90th percentile on our measure predicts that a startup will be 24% less likely to grow by year 1, but this changes to a postive 14% by year six, and 30% by year ten of age. The inverted pattern holds for both IPOs and acquisitions.

Finally, third, we consider the extent to which founding strategy is economically relevant. Using out of sample ROC scores to assess the variance explained, we show that a fully interacted model of four measures of founding differentiation[2] predicts 25% of the variation in the receipt of early stage financing, and 21% of the variation in the receipt of equity growth outcomes. Since our measures are naturally incomplete and imperfect, this estimate is best interpreted as a lower bound on the importance of founding stategy in the data. Furthermore, if one takes seriously the argument in Teece et al. (1997) that high technology startups are the setting in which founding strategies should matter the least, this would also also suggest that the extent to which differentiation predicts performance is generally substantial.

Before delving into our machine learning method, we begin by providing an intuition of our approach. To do so, consider how a human analyst may assess the level of strategic differentiation amongst firms. An analyst may do this by listening to companies' statements, where companies tend to emphasize their differences and unique value propositions. Consider Southwest Airlines and Delta Airlines. Southwest's slogan is "Low fares. Nothing to hide. That's TransFarency!". This slogan emphasizes low cost and transparency, which will be particularly appealing to cost-sensitive customers tired of extra fees. Delta instead uses the slogan "World's Most Trusted Airline", which does not focus on low cost but instead on trust and global coverage. Trust and global coverage might not be as valuable for the cost-sensitive travelers to which Southwest caters, but will be for those travelers that seek to get anywhere reliably and on time, and are willing to pay extra to do so.[3]

In this simple comparison, a strategy analyst can quickly and intuitively identify the differ-

[2]We use the differentiation from the five closest public firms, the single closest public firm, the five closest startups of the same cohort, and the single closest startup of the same cohort.

[3]In fact, Southwest has a relatively low rate of on-time arrivals.

ences between these two statements, even for companies in the same industry. These differences in stated statements do not simply reflect differentiation in the product features. The product in this case is ostensibly similar (a flight), and the statements instead capture the value proposition, be it variety (in route coverage, for Delta) or cost-leadership (Southwest). Even for companies whose competitive advantage is not a unique product, but the ability to deliver a lower price, as in Southwest, these features are emphasized in the company's marketing to consumers. Now, if the strategy analyst is given a third company—such as Spirit Airlines, which has the slogan "Less Money, More Go."—they will also recognize a difference in the perceived 'distance' between this new statement and the prior two. Spirit Airlines appears closer to Southwest, since both focus on the importance of low cost, and will therefore impinge on the differentiation of Southwest more than of Delta.

Our approach expands this idea to all companies and to a richer set of company statements. The key question we ask is the following: if an analyst had a consistent source of marketing materials—such as the website—of all airline carriers, or even all companies in the United States, might she be able to map systematically the conceptual 'distance' between the value proposition of one company and all others? Wouldn't a measure of this distance monotonically reflect higher (or lower) strategic differentiation?

This paper contributes to two distinct areas of the strategy literature. The most important contribution is methodological. This paper provides a formal way to measure founding strategic differentiation, building from the tenets of strategy. The new measure we propose is different from prior attempts at formalizing strategy in that we focus specifically on how a firm is occupying a distinct value proposition from its competitors, as reflected in its marketing. While prior work has instead mostly sought to define the nature of what strategy is (Van den Steen 2016, Porter 1996), there are many applications where measurement itself is of fundamental value. We hope that our results provide useful guidance to researchers seeking to measure differences in the level of strategic positioning and startup founding differentiation. We suspect follow-on work will im-

prove upon our approach. To support this effort, we have released all our code as open source.[4]

Our second contribution is to the conceptual discussion of the role of founding strategy in startup performance. Our results provide empirical insights for broader discussions of whether and how founding differentiation matters in entrepreneurial strategy. In contrast to a view that focuses only on the importance of dynamic capabilities (Eisenhardt & Martin 2000, Teece et al. 1997, McGrath & MacMillan 2000), or experimentation specifically (Kerr et al. 2014, Koning et al. 2019, Gambardella et al. 2018, Reis 2011), we show that founding positioning plays a meaningful role in predicting startup performance. Follow-on work should continue considering the interplay of both founding choices and follow-on dynamic changes. Consistent with a more traditional view of rugged landscapes (Siggelkow 2001), our results suggest that it is both the starting position and the follow-on evolution that determine firm success.

Perhaps the paper closest to ours is the influential work of Hoberg & Phillips (2016) (hereafter HP16). HP16 use the text in the business description section of the annual reports of public firms to develop a new approach to understanding industries and industry dynamics. Specifically, HP16 use the cosine similarity between word vectors weighted by the term frequency inverse frequency algorithm (tf-idf) to estimate a text-based distance between firm statements. Then they implement clustering algorithms to create industry classifications, and show that these perform significantly better than SIC codes. Relative to their work, our paper offers several novel contributions.

First, conceptually, our paper's focus is on strategy rather than industry. This means that while HP16 focus on how companies agglomerate into groups of related firms, strategy focuses on what makes a company distinct from all potential competitors. Understanding those business-specific elements that drive firm performance beyond industry is at the core of strategy research (Rumelt 1991, McGahan & Porter 1997, Ruefli & Wiggins 2003). The precise construct we measure, strategic differentiation, is not measured in any of the papers by HP16, nor is our analysis of how these measures predict either startup profitability or our estimates of the economic impor-

[4]Our code can be found at https://github.com/jorgeguzmanecon/measuring-founding-strategy.

tance of founding strategy for technology-based startups.

Second, methodologically, all our analyses include fixed effects for the text-based industries defined by HP16—created by replicating their methodology within our data—and our regressions cluster our standard errors by these industries. This means that all our effects are effectively estimated conditional on the HP16 industries.

Third, also methodologically, our machine learning model implements a more sophisticated natural language processing method than HP16. Our algorithm uses word embeddings (doc2vec), while HP16 used cosine similarity of relative frequency (tf-idf). The key difference between the two is that the word embeddings approach also incorporates the context in which words are used into the measure. This improvement turns out to be important. Consistent with the idea in strategy that successful positioning requires not only using novel elements, but also combining them in unique ways, we find that the cosine similarity model has a very neglible role in predicting equity outcomes, accounting for only 4% of all variation, while the word embeddings model, in contrast, accounts for 21%. This implies that our measure is substantially more meaningful in an economic sense.

Finally, fourth, from a data perspective, our paper is also the first to focus on startups and to do so using their founding websites. Because websites are usually marketing tools, while 10K statements are investor oriented, our approach is better able to capture market differentiation in itself, while HP16 captured the CEO's perspective on the competitive landscape. We suspect that our approach (and open source code) will be used by other researchers to study other elements of strategy in entrepreneurship and elsewhere.

The rest of the paper proceeds as follows. Section 2 presents our formalized methodological approach. Section 3 reviews our data. Section 4 presents our results. Finally, Section 5 concludes.

2 Measuring founding strategy: a text-based approach

Our measurement approach builds on the idea that written public marketing statements by firms partially reflect their market positioning. A pioneering application using written statements to un-

derstand competitive overlap is found in Hoberg & Phillips (2016).[5] In this section, we introduce a theoretical setup to define our concept more clearly, and its relationship to peformance. Then, we explain how we use firm statements to assess overlap in the value propositions of firms, and how to translate this to a measure of distance. Finally, we explain how we aggregate measures of distance into a specific measure of strategic differentiation, allowing us to score startup differentiation at founding.

Setup

Consider a startup i with some value proposition in the market.[6] At the time it is founded, there are J incumbents, indexed by j, already present in the market. The value proposition of each startup and incumbent is different, but may be related. The consumer has an elasticity of substitution across value propositions ε_{ij} that reflects how their preferences vary between the startup and the incumbents. These elasticities of substitution are aggregated through some function g to into market power M_i, for the startup.

$$M_i = g(\varepsilon_{i1}, ..., \varepsilon_{iJ}) \tag{1}$$

The firm realizes a performance outcome (such as profit) based on this market power, the underlying demand for its product or service, D_i, and a random term μ_i.

$$\Pi_i = h(M_i, D_i)\mu_i \tag{2}$$

Higher expected market power is equivalent to sustained competitive advantage, the goal of a good founding strategy.

Definition 1 (Differentiation score)**.** *For any startup, a measure of their market differentiation* $S_i > 0$ *is a scalar measure that can be positive and monotonically translated to higher market*

[5] See also Abrahamson & Hambrick (1997) for an earlier study relating the language used by firms to industry differences.

[6] This value proposition depends on the startup's product (or service), the startup's cost structure and price, and the way in which the startup delivers this product.

power through some positively increasing function ζ.

$$M_i = \zeta(S_i)\ ,\ \frac{\partial \zeta}{\partial S_i} > 0\ \forall S_i \tag{3}$$

The purpose of our approach is to develop a data-driven way to estimate an empirical equivalent to S_i.

Measuring market differentiation through firm statements

Our approach to measuring market differentiation builds on four important insights. First, while it is virtually impossible to observe the value a consumer sees in a product, it is possible to observe what the firm believes its value proposition to be. In fact, firms constantly state their value proposition with the goal of explaining to consumers (or to some representative set of them) why their product or service should be purchased. Second, the similarity in these marketing statements is a good indicator of the substitutability between the value proposition of each of the products, thus allowing the assessment of how unique (or distant) from other products a new startup is. Third, measuring distance among company statements is not merely a theoretical idea: there are standard text-analysis algorithms that allow us to quantify the relatedness of those statements to effectively create a measure of similarity in the stated value proposition of firms in the market. Distance is then simply the inverse of similarity. Fourth, observing at least some of these statements at or close to founding is possible through the use of archival websites. Since websites represent a de-facto marketing channel for virtually all firms founded after a certain date, the distance between founding websites is a measure of founding positioning.

Building on these insights, we define a new measure of market relatedness σ_{ij}, representing the similarity between two firm statements. Given a startup and an incumbent statements s_i and s_j (one each) explaining their main value proposition, there exists some function h defined between 0 and 1 that can measure a pair-wise similarlity between these two statements as

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

Companies with a value of similarity equal to 1 have completely equivalent statements, while companies with a similarity value of 0 have no relationship to each other. Companies with partial similarity are in between.

Implementing word embeddings

To define h, we focus on a specific implementation of natural language processing (NLP) called word embeddings (word2vec) (Le & Mikolov 2014). Compared to traditional bag-of-words approaches such as the term frequency inverse document frequency algorithm (tf-idf) or topic modeling, word embeddings perform better by allowing the context in which words are used to be incorporated into the algorithm.[7] In essence, while in tf-idf a word is weighted only by how common it is across documents, word2vec represents each word through a vector of N factors that represent the word's characteristics based on the way other words tend to show up around it. These vectors can then be used to understand how similar or different words are to each other. To consider the importance of the document as a whole in our setting, we use doc2vec, which also allows taggings at the document level. h can then be defined directly as the similarity across the vector representations of all the words of each document.[8]

We assume this estimated similarity σ_{ij} is a monotonic (though noisy) representation of the elasticity of substitution between value propositions, and propose that companies with a higher level of similarity in their statements are also more likely to have a higher elasticity of substitution between their value propositions.

$$\frac{\partial E[\varepsilon_{ij}|\sigma_{ij}]}{\partial \sigma_{ij}} > 0 \tag{4}$$

[7]Bag-of-words approaches also allow incorporating semantics by increasing the number of words in each n-gram, however, this tends to quickly lead to sparsity.

[8]Our implementation uses the default parameters from the gensim.models.doc2vec Python library. Specifically, we allow 100 features per word vector, using a distributed memory algorithm (PV-DM), no corpus file, a word window of 5, and 10 total iterations over the corpus (epochs). All code is available at https://github.com/jorgeguzmanecon/measuring-founding-strategy.

From similarity to founding strategy

The next step is to aggregate pair-wise similarity between all startups i and incumbents j into a firm-level strategy score. Our goal is to capture how different the value proposition is between a startup and the incumbent market structure around founding. We first define distance, δ_{ij}, by algebraically inverting σ_{ij}. Distance is a value between 0 and 1, where 0 indicates that two companies are exactly the same and 1 means they are completely different.

$$\delta_{ij} = 1 - \sigma_{ij} \tag{5}$$

Next, we aggregate distance across all incumbents to get an empirical measure of the differentiation score at founding. The mean or median are not good ways to aggregate measures of competitive overlap because most companies are unrelated to each other. Empirical studies in industrial organization highlight how the dynamics of competition are influenced by a small number of competitors and how, as this number increases, the ability of firms to charge margins quickly decreases, approximating a fully competitive economy (in strategy parlance, they lose their competitive advantage). We follow a simple heuristic and use the classic finding of Bresnahan & Reiss (1991) showing that markets become competitive after the first three to five competitors.[9] While this heuristic is admittedly ad-hoc and imperfect, it allows a tractable approach that is applicable across many firms.[10]

The differentiation score can then be estimated by:

$$\hat{S}_i = \frac{1}{5} \sum_{j \in J_i^5} \delta_{ij} \ , \ J_i^5 = \{5\, closest\, incumbents\} \tag{6}$$

We also report analyses focusing on a startup's differentiation from the single closest incumbent, the five closest startups with the same founding year, and the single closest startup in that year.

[9] In more recent, Igami & Uetake (2019) also study the impact of competition on incentives to innovative and find that incentives to innovative drop quickly, stabilizing after five competitors.

[10] In the empirical section, we show that this is not sensitive to the specific number of five competitors, or to whether the averaging approach weights each incumbent by its market value during the year of founding.

3 Data: Crunchbase, the Wayback Machine, and industry controls

We implement this approach on a comprehensive list of startups from Crunchbase, which we complement with their historical websites at the time of founding retrieved through the Wayback Machine, and the annual websites of publicly listed firms in the United States. We use these websites to estimate the founding differentiation scores and the industry of each startup, estimated by replicating Hoberg & Phillips (2016) within our data.

Startup data. We obtained data on all companies available in Crunchbase founded between 2003 and 2019 that have a website and have received some form of financing. Crunchbase is a popular crowd-sourced startup data platform tracking a large number of technology-based startup companies. It is one of the main databases used in entrepreneurship and strategy research, and performs particularly well in covering innovative firms that receive some form of institutional financing (Dalle et al. (2017) provide an overall assessment and examples of the use of Crunchbase in management and economics research). For each company, we downloaded in April 2019 the company name, the founding date, the website address, the city and state of the main office, the date and amount of each financing round, whether the company achieved an equity event (IPO or an acquisition), the market valuation of the company at the exit event, the timing of the exit event, and the top level Crunchbase category for this firm.

Website history data. We used the Wayback Machine, an online platform offered by the Internet Archive (archive.org), to download the initial website of each startup around the time of founding. The Wayback Machine provides access to a digital library containing over 330 billion web-page snapshots occurring in history. These snapshots are taken at least a few times a year for all unique domain names on the internet. We developed a web-scraping technology, available in our Github repository, to automatically query the Wayback Machine for the earliest version of the webpage in the year after the year of founding in Crunchbase. We downloaded the homepage and the first level links in the webpage (up to ten URLs to limit the size of the download). We

excluded all pages that returned empty, that included too little text, or that constituted an HTTP error such as a 403 or 303. Our final dataset includes the founding websites of 13,983 startups.

Incumbent information. To consider the existing market structure at founding we focus on publicly listed firms. Specifically, using the IPO and de-listing dates in Preqin, we downloaded the first available website each year for all companies publicly listed in NASDAQ and the New York Stock Exchange using the same download algorithm used for the startups. This allows us to observe the market proposition of all public companies as stated at the time of startup founding and thus assess adequately the startup's positioning in the market at this time.

Industry Controls. Finally, we develop industry categorization by implementing the method of Hoberg & Phillips (2016) (hereafter, HP16) to develop clusters of related industries based on the company's own business descriptions within our data. To review, HP16 use the term frequency inverse document frequency (tf-idf) algorithm in the business description of 10K annual reports to develop vectors of weighted words, and then the cosine similarity between these vectors to estimate a scalar distance from one company to another. Then, they implement a k-means clustering algorithm and use the resulting cluster identifiers as the industry categorization. HP16 recommend using 300 clusters as the target number to mimic well the distribution of SIC industries.[11] We implement this method with 300 industries using our website text rather than the 10K business descriptions, to define industries within our data. The resulting variable *HP Industries*, represent 300 indicators for the clusters created through this method. The median number of startups in an industry is 36, and the average 50.

Summary statistics

Table 1 presents summary statistics of our data. The average founding website length for a startup is about thirteen thousand characters, but this measure is significantly skewed. 67% of firms in the Crunchbase recieve early stage financing, with an average value of $869 thousand. 31% get series A financing, with an average of $2.6 million. 17% of firms achieve an equity growth out-

[11] This algorithm is the fixed-industry classification algorithm in Hoberg & Phillips (2016) (p. 1435). Hoberg & Phillips also implement a 'network' based measure which our approach does not allow us to implement.

come, of which IPO represents only 1.8% and acquisition is 15%.

4 Results

We now proceed to report our empirical estimates. We begin by reporting estimates of founding differentiation relative to public firms and other startups, as well as providing examples of our measure across specific industries. Our measure of founding differentiation appears to relate well to usual conceptualizations of this construct. Then, we perform three empirical explorations with this measure. First, we study the statistical correlation between our differentiation score and early stage financing, both unconditionally and conditional on a series of fixed effects. The relationship of our main measure of founding differentiation is positive and significant, suggesting that being different from the incumbent market structure (proxied by public firms) predicts early stage performance. In contrast, differentiation of the focal startup from other startups, who are competitors in the startup financing market, is much noisier and either negative or not significant (depending on the specification). We interpret these relationships as consistent with our measure capturing strategic differentiation in the consumer market, rather than in the financing market. Next, we study the extent to which our differentiation scores predict long-term outcomes such as equity exit events, such as IPO and acquisitions. We again find a positive and significant relationship but in this case it is much weaker and appears significant at the usual levels only for certain subsamples or subsets of outcomes. While this initially appears puzzling, it becomes clearer when we study the dynamics of our effect. Indeed, the strategy literature has highlighted how differentiated companies may face additional challenges early on to achieve legitimacy, but nonethess perform better over time (Deephouse 1999, Marx et al. 2014). This is exactly the pattern we observe. While more differentiated firms initially are less likely to exit, this relationship inverts over the lifecycle, for both IPOs and acquisitions. Once again, we find the strength of all these relationships is greater for measures of differentiation from public firms, and not other startups. Finally, we move to study the degree to which our measure accounts for variation in firm performance. To do so, we use all our text-based measures to plot out of sample ROC scores for both getting early

stage financing and equity outcomes. Our measures account for 25% and 21% of variation in outcomes, respectively. We interpret these as lower-bound estimates of the economic importance of founding strategy on startup performance.

Estimates of similarity and founding strategy

Table 2 reports summary statistics for our four estimated differentiation scores. Our preferred measure is *Differentiation Score (5 Closest Public Firms)*, defined as the average distance from the focal startup's website at founding to the five closest public firms. It has a mean value of 0.4 and a standard deviation of 0.11. The difference between the 10th and 90th percentile is 0.274. We also include measures for the distance from the single closest public firm, the average distance from the five closest Crunchbase startups founded in the same year, and the distance from the closest Crunchbase startup founded in the same year. Figure 1 also reports the histogram of this measure, showing that the distribution is almost normal. Table 3 reports the correlation between these four measures. All measures are highly correlated, though the measures of public firms are naturally more correlated to other public firms, and similarly for startups.

Next, we provide, in Tables 4 and 5, examples of how our measure scores individual companies within an industry. Table 4 reports the differentiation score of a select set of startups categorized in Crunchbase as belonging to the food industry, and the startup's description as written in Crunchbase.[12] Consistent with the idea that we are able to measure strategic differentiation, we observe at the top of the distribution firms that describe substantively new and strongly differentiated products, such as jicama chips (JicaChips) or the 'blockchain of food' (Ripe.io). In the middle, we observe companies that appear more traditional, though still focusing on a relatively novel niche, such as a seaweed-focused food (Ocean's Halo). Finally, at the bottom of the distribution, we see firms who have very traditional products, such as retail discounting (Big Box Overstocks).

Table 5 repeats the same exercise in consumer electronics. At the top of the distribution, we

[12]It is useful to note that this description is independent of the website text we use to develop our measurse, therefore serving as an additional point of verification.

observe significantly differentiated companies, focused on creating wifi-enabled teddy bears (Parihug) and AI to manage chronic disease (Lark). The middle appears more traditional, with applications such as music sharing (Ringz.TV) and remote control for presentations (Penxy). Finally, the bottom focuses on companies that provide what may be typical goods such as warranty management. Tables A1 and A2 provide a longer list of firms in the food and consumer electronics industry.

Founding strategy score and early stage financing outcomes

We next assess the predictive relationship between our differentiation score and early stage financing outcomes. Figure 2 presents binned scatterplots elucidating these relationships. Panel A reports a strong unconditional correlation within firm cohorts. A higher differentiation score predicts a higher amount of early stage financing raised. Panel B replicates what will become our preferred specification, introducing both founding year fixed effects, and fixed effects for HP industries. The pattern is slightly less pronounced, but still positive and meaningful. Panel C considers the possibility that there is something about the way the websites are built that correlates to both our measure and outcomes, by controlling for the length of website text. To do so, we split our variable of website text length into 20 bins and include them as additional fixed effects. Reassuringly, our correlation looks similar. Finally, Panel D reports the extensive margin, a binary measure of whether a startup gets financing or not, with a similarly positive result.

We present additional scatterplots using Series A financing events rather than early stage events in Appendix Figure A1, and our alternative measures of differentiation score in Appendix Figure A2. The key relationships do not change using Series A financing events. Further, we observe a similar positive correlation for our strategic differentiation score using the single closest public firm. In contrast, the differentiation score estimated from distance to other startups appears to be much less correlated to financing.

Main Relationships. We consider these relationships in more detail in Tables 6 and 7. These tables report a regression analysis studying how founding diffentiation score predicts perfor-

mance. In Table 6, we study the extensive margin of financing by running an OLS regression with *Gets Early Stage Financing* as the dependent variable. Standard errors are double clustered by HP industry and state to account for industry or location correlation among observations. Column (1) shows a positive unconditional coefficient of 0.68. Column (2) shows that there are large cohort effects: the coefficient drops to 0.16 after including founding year fixed effects. Column (3) is our preferred specification, which includes both founding year and HP industry effects. The coefficient is 0.06.[13] To put this result into perspective, this implies that moving from the 10th to the 90th percentile of our measure predicts a 1.7 percentage points (about 3% of the mean) higher likelihood of raising early stage financing. Finally, Column (4) is an additional robustness test that includes city fixed effects and state by year fixed effects to account for the possibility of geographic unobservables driving our effect. Our coefficient is unchanged.

Next, in Table 7, we study the total amount of financing received, by using *Log(Early Stage+1)* as the dependent variable. The differences here are more dramatic. Columns (1) and (2) show that there are similarly large cohort effects in our data. The coefficient for our preferred specification, is column (3), using founding year and HP industry fixed effects. The estimate is 0.82. This implies that moving from the 10th to the 90th percentile predicts a 25% increase in total early stage fundraising,[14] and moving by one standard deviation predicts a 9.5% increase in early stage financing. Column (4) shows that this is robust to including geography based fixed effects. We incorporate two robustness tests in columns (5) and (6). Column (5) uses the series A financing events instead of early stage financing, and column (6) includes both early stage financing and series A together. The coefficients are larger, 1.47 and 1.32, respectively. These imply that moving from the 10th to 90th percentile increases series A financing by 49% and the total of early stage and series A financing by 44%.

Together, these relationships document a strong and positive predictive relationship between our estimate founding differentiation score and the receipt of early stage financing, validating

[13]This drop of 62% is lower than the classic estimate in Schmalensee (1985) on the role of industry to firm profitabilty, but higher than the follow-on estimates in Rumelt (1991) and McGahan & Porter (1997).

[14]The 10th-90th percentile range is 0.274; $e^{(.274*0.82)} - 1 = 0.25$.

the ability of our measure to capture founding differentiation, as well as serving as comparative statics on this relationship.

Other Differentiation Scores. Table 8 next considers the relationship of other differentiation scores to early stage financing. Column (1) repeats the preferred estimate of Table 7 for comparability, using the differentiation score estimated from the 5 closest public firms. Column (2) instead uses the differentiation from the single closest firm. The coefficient is very similar, though slightly larger at 1.020 and the effect slightly more precise. Columns (3) and (4) next consider a different type of differentiation, differentiation from other startups in the same cohort. While differentiation from public firms is intended to capture the market positioning in the extant U.S. economy, differentiation from other startups may better reflect the uniqueness and positioning in the venture financing market or other startup resourcs. Interestingly, these coefficients are negative and sometimes significant. The differentiation from the 5 closest startups has a coefficient of -0.85 and is significant, while the differentiation from the closest startup has a coefficient of -0.14 and is not significant. These differences are consistent with our view that what we are measuring is the startup's overall positioning in the product market itself, and that this subsequently determines performance.

Columns (5), (6), and (7) introduce multiple measures at the same time to consider the correlation of one measure of differentiation conditional on others. The patterns appear basically the same. We observe a positive coefficient for the differentiation from public firms. In contrast, differentiation from other startups appears, if anything, negative.

These results are further evidence of the ability of our measures to capture well the market positioning of startups and the way in which our measure of positioning (the differentiation score) predicts performance.

Founding differentiation score and equity outcomes

We next study whether our differentiation score predicts long term firm outcomes by examining its relationship to equity success through IPO or acquisition. Figure 3 reports binscatters of our differentiation score with *Equity Growth*, a binary measure equal to 1 if the firm achieves IPO or

acquisition, as the dependent variable. Panel A shows a strong unconditional correlation between both measures. Panel B shows the correlation becomes even stronger once fixed effects for HP industries are included. Panel C shows that this correlation remains once website length is controlled for.

Main Relationships. We proceed to study these relationship in more detail through regressions in Table 9. Standard errors are double clustered by HP industry and state to account for correlation across industry or geography. The relationships appear positive, though only sometimes significant at the usual levels. Column (1) is an unconditional correlation, which is positive and not significant. In column (2) we control for the growth rates of different industries by including the leave-one-out mean growth as a control in the regression. This value is the average growth for firms in the industry that are not the focal firm. The coefficient is now positive and significant at 0.046. This implies that moving from the 10th to 90th percentile results in a 1.3 percentage points increase in the probability of growth, or 7.5% of the mean. Column (3) includes the fixed effects for HP industries. The coefficient is still positive but is now not significant at the usual levels (p=0.16). Column (4) controls for state and city fixed effects, and columns (5) and (6) look at only IPOs and only acquisitions. These estimates are again positive, but noisy.

In Table 10 we perform additional analyses to provide more detail on the relationship of our variables to equity performance. Column (1) drops all firms that are young in the data, and for whom consequently we cannot observe their whole lifecycle. To do so, we remove all firms born in 2016 or later. The coefficient is now positive and significant, with a value of 0.048. This result foreshadows the possibility of a dynamic effect on the way differentiation predicts equity outcomes: the positive relationship between founding differentiation and equity outcomes appears to be missing for very young firms, and to only exist when considering the whole firm lifecycle. We will explore these dynamics in more detail in Table 12. Column (2) drops from the outcome variable all low value acquisitions, which we define as either having no sale price in Crunchbase or having a sale price under 10 million dollars. Again, the coefficient is now more precisely estimated and significant. This suggests that the positive relationship we focus on is

also more relevant for meaningful equity outcomes, rather than fire-sale. Column (3) controls for the initial amount of financing to study whether the relationship to performance holds conditional on the early success we documented previously. The coefficient is positive and slightly noisy (p=0.13). Finally, column (4) uses the price of acquisition within acquired firms, to study the extent to which it changes with founding differentiation. The coefficient is positive but not significant.

Table 11 studies whether our other differentiation scores also predict equity growth. While all coefficients are positive, no measure is significant at the usual levels.

These results emphasize a positive relationship of founding strategy to performance, but one that appears noisy on the aggregate. It becomes more robust when considering only firms that are observed throughout their lifecycle, and more meaningful acquisitions. The next section studies these dynamics.

Dynamics. So far, we have limited ourselves to cross sectional regressions relating founding differentiation to outcomes. In the case of equity outcomes, this appears unsatisfactory given that existing literature suggests the dynamics of acquisitions and IPOs may vary depending on how differentiated a firm is at founding. One particularly natural hypothesis is that highly differentiated startups may take longer to develop as they face additional legitimacy costs in the market (Deephouse 1999). Indeed, Marx et al. (2014) show that firms that have disruptive technologies (which tend to be more differentiated products) develop dynamic commercialization strategies that require the firm to compete to eventually get to acquisition outcomes. In this case, for example, the less disruptive companies are *more* likely to be acquired early on, but this pattern changes as the firm ages and disruptive companies achieve necessary legitimacy. For our analysis, this type of dynamics would create a negative bias, since many firms are only observed for a few years (e.g., the younger cohorts) and we do not get to observe the later years in the life cycle.

We address this possibility in Table 12. To do so, we estimate our preferred specification in a panel format and report the individual coefficients for founding differentiation score against the cumulative probability that a firm has observed an equity exit by each year of age. The pattern

we see is dramatic. Firms with a higher differentiation score are initially less likely to exit, particularly during their first year (year 0). But this pattern reverses as the firm ages. For our main growth outcome, the coefficient turns positive by age 5 and continues increasing thereafter. By age 7, firms with a higher differentiation score are much more likely to have had an equity exit. The coefficient, with a value of 0.126, implies that moving from the 10th to the 90th percentile of our measure is associated with a 3.4 percentage points higher likelihood of exit, or 21% of the mean, and an increase by one standard deviation with 1.3 percentage points (8% of the mean). Columns (2) and (3) show that these patterns hold for both IPOs and acquisitions independently. For IPOs, the coefficient is significant by age 7, with a value of 0.018. Since only 1.7% of the firms in the sample reach an IPO, this effect is quite meaningful. The coefficient for acquisition is indeed higher, turning positive and significant by age 5, and ultimately being even higher than the overall sample mean of acquisitions. By age 10, the positive relationship to founding differentiation is sbustantial. A startup scoring at the 90th versus the 10th percentile of our measure is predicted to be 26% more likely to IPO and 31% more likely to be acquired. Importantly, we emphasize that our outcomes variable is cumulative, so that the positive effect reflects the total success up to that year, including the negative impact of the early years.

Together, these results suggest important dynamics for equity outcomes that are consistent with our differentiation score ultimately capturing the nature of more innovative or unique ideas. These ideas do tend to perform better over the long term, but they may take a longer time to do so, possibly due to the cost of educating and better understanding the market to either acquire legitimacy (Deephouse 1999) or to prove technological feasibility (Marx et al. 2014).

How much does founding strategy matter?

Finally, we study the extent to which our measures, and consequently founding strategy are economically relevant. To do so, we perform an out of sample analysis to consider how much variation in outcomes can be predicted by our measures. Specifically, we use a 10-fold approach to regress a logit model on a fully interacted version of our four founding differentiation scores (5 closest public firms, closest public firm, 5 closest cohort startups, and closest cohort startup) on

the binary version of both of our outcomes, *Gets Early Stage Financing* and *Equity Growth*. We then store the out of sample predictions from these models,[15] Ond consider how well do they predict actual outcomes. The results are reported in Figure 4.

Panels A and B consider early stage financing. Panel A reports the share of firms that receive early stage financing across the distribution of out of sample predicted probability. We observe a positive and increasing slope, with firms at the top end of the distribution being almost twice as likely to get early stage financing as firms at the bottom. Panel B is our preferred measure. It reports the out of sample ROC score (area under the curve) of this model. This is the preferred approach to assess the predictive fit of binary models. The ROC score conceptually answers the following question: if two firms, one with early stage financing and one without, are fed to the model, what is the likelihood that the one with early stage financing is scored higher by the model? A random model would have an ROC of 0.5, and a fully informative one would be 1. The graph shows an ROC value of 0.626, implying the model is able to account for about 25% of the variation in outcomes.

Panels C and D consider the same two statistics for the equity growth outcome. Once again we observe a meaningful ability of our index to predict performance. Firms in the top ventiles of the out of sample predicted index are more than three times more likely to have an equity growth outcome than firms in the bottom ventiles. The ROC score is slightly lower at 0.606, implying that our model is able to account for 21% of variation in outcomes.

Appendix Figure A3 repeats the model using a simpler measure of similarly—the cosine of the term frequency inverse document frequency (tf-idf) estimates. Interestingly, while these measures account for about the same amount of variation in early stage financing, the ROC score for the equity growth outcome is only 0.523, implying the tf-idf measures only account for about 4% of total variation. Finally, Appendix Figure A4 reports a model that uses both HP industries and differentiation scores together. The ROC scores increase, but only moderately, to 0.67 for *Gets Venture Capital* and 0.636 for *Equity Growth*.

[15] In essence, we split the data into 10 random subsamples, and for each subsample, we use the predicted value from a regression using all other 9 subsamples but excluding the focal one.

Together, these estimates provide a novel assessment for the importance of founding positioning on overall performance. We show that founding positioning accounts for 25% and 21% of variation in outcomes out of sample, using our measures. Drawing a parallel to with the theoretical concept that it is not only using unique resources, but being able to use them together in novel ways, that creates an interesting strategic positioning, the tf-idf model scores much lower. Furthermore, because our measures are inherently noisy, our estimate on the economic importance of founding positioning is a lower bound within in our sample.

5 Conclusion

Founding strategy is the set of characteristics present during (or around) startup founding that relate to sustained competitive advantage. Building from the key ideas of strategic positioning, we have introduced a systematic approach that uses natural language processing tools to measure the differentiation of firms, and applied it to Crunchbase data to estimate the founding differentiation of a large number of technology based startups. Our approach takes advantage of the key insight that a large portion of firms state their value proposition close to founding in their website. We focus on the historical websites of the firms and the text they wrote around the time of founding. Using NLP techniques that take into account both the incidence and context in which words are used (word embeddings), we develop a measure of similarity between all websites from individual cohorts and estimate strategic differentiation as the mean distance from the closest incumbent firms (proxied by public firms). After validating our measure qualitatively through examples, we perform analyses that show that this measure predicts follow-on financing and equity outcomes, and accounts for up to 25% of their variance, suggesting that founding differentiation is economically important. This approach is replicable and is accompanied by open source Python code that allows any other researcher to also download all historical websites and to re-create our founding strategy estimates, or expand their use outside of entrepreneurship.

The goal of our paper has been to make three unique contributions. First, and most importantly, we have sought to connect the use of machine learning techniques to the understanding of

strategy. The relationship between the two is obvious: machine learning is a tool to solve a prediction problem (Agrawal et al. 2018), and sustained competitive advantage, the goal of a good strategy, is precisely higher predicted profits for firms. Therefore, a mapping of firm conditions to performance should in principle advance the understanding and measurement of strategy. We applied specific machine learning techniques within the conceptual tenets of one idea in strategy (strategic positioning) and created a useful measure that can substantially inform research and practice. We hope our foray is only the beginning of a deeper relationship between machine learning and strategy, to understand the determinants of firm performance.

Second, we have also more directly sought to study startup founding strategy and assess the extent to which it can be measured. Whether startup founding strategy can be measured is not obvious. Influential work such as Lean Startup, for example, posits that startup founding conditions do not matter (Reis 2011) but instead only startups' evolution. Our approach shows that one component of founding strategy, market differentiation, can be measured using widely available tools, and that this measurement has meaningful relationships to startup outcomes and their dynamics. We hope that our specific differentiation measurements can be incorporated into entrepreneurship research and contribute to understanding the dynamics of startup performance.

Finally, to this end, we also provide specific estimates on the relationship between founding differentiation and startup performance. We emphasize these these estimates are not causal, but comparative statics that formalize our understanding of the cross-sectional relationship between the two. We show that founding differentiation is important only when considered vis-á-vis the incumbent market structure, and not other startups. It holds a clear and strong relationship to early stage financing outcomes, such as seed funding or angel funding, and to series A financing. It also correlates positively to equity outcomes, but the relationship is more nuanced: founding differentiation predicts negative outcomes in the early years, but it predicts total positive outcomes after age five. In alignment with prior work, this may suggest that more differentiated firms may take longer to establish themselves, but may eventually be more successful (Marx et al. 2014). We also provide original estimates on the economic importance of founding differentia-

tion for performance, and showing that founding differentation predicts 25% of the out of sample variation in early stage financing, and 21% of the variation in equity outcomes. Highlighting the value of our specific word embeddings approach in characterizing firms, using a simpler NLP method that does not take word context into account only characterizes 4% of the variance in equity outcomes.

References

Abrahamson, E. & Hambrick, D. C. (1997), 'Attentional homogeneity in industries: the effect of discretion', *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* **18**(S1), 513–532.

Aghion, P. & Howitt, P. (1992), 'A model of growth through creative destruction', *Econometrica: Journal of the Econometric Society* pp. 323–351.

Agrawal, A., Gans, J. & Goldfarb, A. (2018), *Prediction machines: the simple economics of artificial intelligence*, Harvard Business Press.

Anthony, C., Nelson, A. J. & Tripsas, M. (2016), '"who are you?... i really wanna know": Product meaning and competitive positioning in the nascent synthesizer industry', *Strategy Science* **1**(3), 163–183.

Barney, J. (1991), 'Firm resources and sustained competitive advantage', *Journal of management* **17**(1), 99–120.

Baumol, W. J. (1968), 'Entrepreneurship in economic theory', *The American economic review* **58**(2), 64–71.

Bresnahan, T. F. & Reiss, P. C. (1991), 'Entry and competition in concentrated markets', *Journal of political economy* **99**(5), 977–1009.

Dalle, J.-M., den Besten, M. & Menon, C. (2017), Using crunchbase for economic and managerial research, Working paper, OECD Science, Technology and Industry Working Papers.

Deephouse, D. L. (1999), 'To be different, or to be the same? it'sa question (and theory) of strategic balance', *Strategic management journal* **20**(2), 147–166.

Eisenhardt, K. M. & Martin, J. A. (2000), 'Dynamic capabilities: what are they?', *Strategic management journal* **21**(10-11), 1105–1121.

Gambardella, A., Camuffo, A., Cordova, A. & Spina, C. (2018), 'A scientific approach to entrepreneurial decision making: Evidence form a randomized control trial', *Management Science, Forthcoming* .

Gans, J., Scott, E. & Stern, S. (2019), *Entrpreneurial Strategy*, Manuscript.

Guzman, J. & Stern, S. (2019), 'The state of american entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 us states, 1988-2014', *American Economic Journal: Economic Policy* .

Hannan, M. T. & Freeman, J. (1977), 'The population ecology of organizations', *American journal of sociology* **82**(5), 929–964.

Hellmann, T. & Puri, M. (2000), 'The interaction between product market and financing strategy: The role of venture capital', *The review of financial studies* **13**(4), 959–984.

Hoberg, G. & Phillips, G. (2016), 'Text-based network industries and endogenous product differentiation', *Journal of Political Economy* **124**(5), 1423–1465.

Igami, M. & Uetake, K. (2019), 'Mergers, innovation, and entry-exit dynamics: Consolidation of the hard disk drive industry, 1996-2016', *Available at SSRN 2585840* .

Kaplan, S. N., Sensoy, B. A. & Strömberg, P. (2009), 'Should investors bet on the jockey or the horse? evidence from the evolution of firms from early business plans to public companies', *The Journal of Finance* **64**(1), 75–115.

Kerr, W. R., Nanda, R. & Rhodes-Kropf, M. (2014), 'Entrepreneurship as experimentation', *Journal of Economic Perspectives* **28**(3), 25–48.

Knight, F. H. (1921), 'Risk, uncertainty and profit'.

Koning, R., Hasan, S. & Chatterji, A. (2019), Experimentation and startup performance: Evidence from a/b testing, Working Paper 26278, National Bureau of Economic Research. **URL:** *http://www.nber.org/papers/w26278*

Le, Q. & Mikolov, T. (2014), Distributed representations of sentences and documents, *in* 'International conference on machine learning', PMLR, pp. 1188–1196.

Lerner, J. & Schoar, A. (2010), 'Introduction to" international differences in entrepreneurship"', pp. 1–13.

Marx, M., Gans, J. S. & Hsu, D. H. (2014), 'Dynamic commercialization strategies for disruptive technologies: Evidence from the speech recognition industry', *Management Science* **60**(12), 3103–3123.

McGahan, A. M. & Porter, M. E. (1997), 'How much does industry matter, really?', *Strategic management journal* **18**(S1), 15–30.

McGrath, R. G. & MacMillan, I. C. (2000), *The entrepreneurial mindset: Strategies for continuously creating opportunity in an age of uncertainty*, Vol. 284, Harvard Business Press.

Moreira, S. (2016), 'Firm dynamics, persistent effects of entry conditions, and business cycles', *Persistent Effects of Entry Conditions, and Business Cycles (October 1, 2016)* .

Nelson, R. R. & Winter, S. G. (2002), 'Evolutionary theorizing in economics', *Journal of economic perspectives* **16**(2), 23–46.

Porter, M. E. (1996), 'What is strategy?', *Harvard Business Review* **6**(74), 61–78.

Reis, E. (2011), *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, Currency.

Ruefli, T. W. & Wiggins, R. R. (2003), 'Industry, corporate, and segment effects and business performance: a non-parametric approach', *Strategic Management Journal* **24**(9), 861–879.

Rumelt, R. P. (1991), 'How much does industry matter?', *Strategic Management Journal* **12**(3), 167–185.
URL: *http://www.jstor.org/stable/2486591*

Schmalensee, R. (1985), 'Do markets differ much?', *The American economic review* **75**(3), 341–351.

Schumpeter, J. (1942), 'Capitalism, socialism and democracy', *New York* .

Shane, S. A. (2003), *A general theory of entrepreneurship: The individual-opportunity nexus*, Edward Elgar Publishing.

Siggelkow, N. (2001), 'Change in the presence of fit: The rise, the fall, and the renaissance of liz claiborne', *Academy of Management journal* **44**(4), 838–857.

Stinchcombe, A. L. (1965), 'Organizations and social structure', *Handbook of organizations* **44**(2), 142–193.

Teece, D. J., Pisano, G. & Shuen, A. (1997), 'Dynamic capabilities and strategic management', *Strategic management journal* **18**(7), 509–533.

Van den Steen, E. (2016), 'A formal theory of strategy', *Management Science* **63**(8), 2616–2636.

Winter, S. G. (2003), 'Understanding dynamic capabilities', *Strategic management journal* **24**(10), 991–995.

Table 1: Summary Statistics Crunchbase Firms

	(1)			
	mean	sd	min	max
Website Text Length	13564.998	22640	301	238515
Early Stage Financing (Thousands $)	868.468	3248	0	191000
Gets Early Stage Financing	0.667	.471	0	1
Series A Financing (Thousands $)	2615.420	8677	0	411770
Gets Series A	0.308	.462	0	1
IPO	0.018	.131	0	1
Acquisition	0.152	.359	0	1
Growth	0.171	.376	0	1
Observations	13983			

Dataset is all companies in Crunchbase founded since 2003 that raised financing and for whom we where able to download a founding website. Founding website is downloaded from the WaybackMachine as the earliest website available the year after founding. Early Stage Financing is defined as all financing that is seed financing, angel financing, or grants.

Table 2: Summary Statistics of Strategic Differentiation Score

	mean	sd	p10	p50	p90	min	max
Differentiation Score (5 Closest Public Firms)	0.397	.105	.25	.4	.53	.074	.78
Differentiation Score (5 Closest Cohort Startups)	0.346	.0907	.22	.35	.46	.069	.64
Differentiation Score (Closest Public Firm)	0.372	.102	.23	.38	.5	.056	.77
Differentiation Score (Closest Cohort Startups)	0.312	.0894	.19	.32	.43	.054	.61
Observations	13983						

Strategic differentiation score represents the conceptual distance in the market between a firm and some of its closest competitors. It is estimated in three steps. First, a measure of similarity is estimated between the founding website of all startups in a cohort and the website of all public firms during the startup year of founding. To do so, we use a word embeddings algorithm that accounts for both the incidence of words and their context. Next, distance is defined as one minus this similarity. Finally, differentiation is the average distance to the closest competitors. We report four measures. The distance to the five closest incumbent firms (public firms). Distance to the single closest public firm. Distance to the five closest startups from the same cohort. And distance to the single closest startup in the same cohort. All code is available at https://github.com/jorgeguzmanecon/measuring-founding-strategy.

Table 3: Correlation of Differentiation Scores

	(1)	(2)	(3)	(4)
(1) Differentiation Score (5 Closest Public Firms)	1.00			
(2) Differentiation Score (Closest Public Firm)	0.98	1.00		
(3) Differentiation Score (5 Closest Cohort Startups)	0.80	0.78	1.00	
(4) Differentiation Score (Closest Cohort Startup)	0.73	0.72	0.95	1.00

Table 4: Examples of Strategic Differentiation Score: Food Industry

Percentile	Diff. Score (5 Closest Public Firms)	Website	Company Name	Short Description
.94	.55	www.jicachips.com	JicaChips	JicaChips- World's 1st Jicama Chip! Less than 100 Calories Per Bag.
.88	.52	ripe.io	Ripe.io	Ripe.io is creating the Blockchain of Food.
.77	.48	www.impactvi.com	ImpactVision	ImpactVision is building a new standard for food safety and quality using hyperspectral technology.
.49	.4	oceanshalo.com	Ocean's Halo	Ocean's Halo is a seaweed-focused food company.
.31	.35	here.co	Here Holdings	Here Holdings creates food and beverage products in Illinois using Midwest produce.
.071	.23	oceanapproved.com	OCEAN APPROVED	OCEAN APPROVED provides domestic, fresh, healthy alternative to imported seaweed products.
.024	.18	www.bigboxoverstocks.com	Big Box Overstocks	Big Box Overstocks is a discounts store that offers discontinued, damaged packaging, end of season, oversupply, and open box items.

We report a selected sample of firms that are categorized in Crunchbase as part of the Food industry. We include their estimated differentiation score, the percentile of this score within the distribution of all firms, the name, the website, and the short description that is included for these companies in Crunchbase. A longer list of companies in this industry is included in Appendix Table A.1.

Table 5: Examples of Strategic Differentiation Score: Consumer Electronics

Percentile	Diff. Score (5 Closest Public Firms)	Website	Company Name	Short Description
.99	.62	parihug.com	Parihug	Parihug makes pairable, wifi-enabled teddy bears that let you hug someone from anywhere in the world.
.92	.53	www.lark.com	Lark	Lark is the leading digital health company using AI and clinical science to deliver scalable, positive health outcomes in chronic disease.
.72	.46	vespermems.com	Vesper	Vesper is a designer of advanced acoustic-sensing technology.
.55	.41	shadecraft.com	ShadeCraft	Improving human life outdoors through robotic technology .
.51	.4	ringz.tv	Ringz.TV	Ringz, an addictive video sharing application, connects users to shared playlists that are watchable on any web-connected device.
.26	.33	penxy.com	Penxy	Penxy is a slide sharing application for presenters to control their presentations in real time via iOS devices.
.083	.24	www.unitedkeys.com	United Keys	United Keys engages in the development of technology for PC display input devices such as keyboards and keypads to private label customers.
.067	.23	www.crosswarranty.com	CrossWorld Warranty	CrossWorld Warranty is a platform for manufacturers and retail consumers to handle their warranty records online.

We report a selected sample of firms that are categorized in Crunchbase as part of the Consumer Electronics industry. We include their estimated differentiation score, the percentile of this score within the distribution of all firms, the name, the website, and the short description that is included for these companies in Crunchbase. A longer list of companies in this industry is included in Appendix Table A.2.

Table 6: Does founding differentiation predict the receipt early stage financing?

	(1)	(2)	(3)	(4)
Differentiation Score (5 Closest Public Firms)	0.678***	0.156***	0.0615***	0.0658**
	(0.0795)	(0.0390)	(0.0166)	(0.0268)
Founding Year F.E.	No	Yes	Yes	No
HP Industry F.E.	No	No	Yes	Yes
Year × State F.E.	No	No	No	Yes
City F.E.	No	No	No	Yes
Observations	13983	13983	13983	13983
R^2	0.023	0.121	0.176	0.276

OLS linear probability model. Dependent variable is equal to 1 if a startup gets early stage financing (seed or angel financing) and zero otherwise. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 7: Does founding differentation predict the amount of early stage financing?

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep. Var. Early Stage	*Dep. Var.* Early Stage	*Dep. Var.* Early Stage	*Dep. Var.* Early Stage	*Dep. Var.* Series A	*Dep. Var.* Series A + Early Stage
Differentiation Score (5 Closest Public Firms)	8.827***	1.850***	0.817***	0.898**	1.465*	1.323***
	(1.064)	(0.467)	(0.237)	(0.381)	(0.792)	(0.417)
Founding Year F.E.	No	Yes	Yes	No	Yes	Yes
HP Industry F.E.	No	No	Yes	Yes	Yes	Yes
Year X State F.E.	No	No	No	Yes	No	No
City F.E.	No	No	No	Yes	No	No
Observations	13983	13983	13983	13983	13983	13983
R^2	0.022	0.115	0.162	0.259	0.129	0.045

OLS model. Dependent variable is the log of total fundraised in early stage financing plus 1. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 8: Other measures of founding differentiation and early stage financing.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Differentiation Score (5 Closest Public Firms)	0.817**				-3.677	3.922***	
	(0.237)				(2.355)	(0.666)	
Differentiation Score (Closest Public Firm)		1.020***			4.682		2.169***
		(0.229)			(2.377)		(0.586)
Differentiation Score (5 Closest Cohort Startups)			-0.852**			-4.425***	
			(0.315)			(0.825)	
Differentiation Score (Closest Cohort Startups)				-0.142			-1.830*
				(0.371)			(0.756)
Founding Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HP Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13983	13983	13983	13983	13983	13983	13983
R^2	0.162	0.163	0.162	0.162	0.163	0.163	0.163

OLS model. Dependent variable is the log of total fundraised in early stage financing plus 1. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 9: Does founding differentiation predict equity performance?

	(1) *Dep. Var.* IPO or Acq.	(2) *Dep. Var.* IPO or Acq.	(3) *Dep. Var.* IPO or Acq.	(4) *Dep. Var.* IPO or Acq.	(5) *Dep. Var.* IPO	(6) *Dep. Var.* Acquisition
Differentiation Score (5 Closest Public Firms)	0.0332	0.0463**	0.0377	0.0318	0.00746	0.0302
	(0.0225)	(0.0221)	(0.0269)	(0.0269)	(0.00956)	(0.0255)
Founding Year F.E.	Yes	Yes	Yes	No	Yes	Yes
HP Industry F.E.	No	No	Yes	Yes	Yes	Yes
Year × State F.E.	No	No	No	Yes	No	No
City F.E.	No	No	No	Yes	No	No
Mean Industry Growth	No	Yes	No	No	No	No
Observations	13983	13983	13983	13983	13983	13983
R^2	0.083	0.092	0.124	0.212	0.109	0.112

OLS model. Dependent variable is a binary variable equal to 1 if a firm is IPO or acquired and zero otherwise. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 10: Founding differentiation and equity performance

	(1) *Subsample:* Drop Firms from 2016 and Later	(2) *Change in D.V.:* Drop Low Valuation Acquisitions	(3) Control for Early Stage Financing	(4) *Dep. Var.* Log(Acq. Price)
Differentiation Score (5 Closest Public Firms)	0.0480*	0.0248**	0.0417	0.956
	(0.0271)	(0.00949)	(0.0268)	(0.767)
Log(Early Stage + 1)			-0.00491**	
			(0.000643)	
Founding Year F.E.	Yes	Yes	Yes	Yes
HP Industry F.E.	Yes	Yes	Yes	Yes
Observations	12575	13983	13983	360
R^2	0.114	0.106	0.129	0.401

OLS model. Dependent variable is a binary variable equal to 1 if a firm is IPO or acquired and zero otherwise. HP Industry fixed effects are fixed effects for 300 industrie created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 11: Other measures of founding differentiation and equity performance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Differentiation Score (5 Closest Public Firms)	0.0377				0.178	-0.00235	
	(0.0269)				(0.129)	(0.0558)	
Differentiation Score (Closest Public Firm)		0.0310			-0.147		0.00989
		(0.0274)			(0.131)		(0.0345)
Differentiation Score (5 Closest Cohort Startups)			0.0549			0.0570	
			(0.0392)			(0.0757)	
Differentiation Score (Closest Cohort Startups)				0.0414			0.0337
				(0.0355)			(0.0462)
Founding Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HP Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13983	13983	13983	13983	13983	13983	13983
R^2	0.124	0.124	0.124	0.124	0.124	0.124	0.124

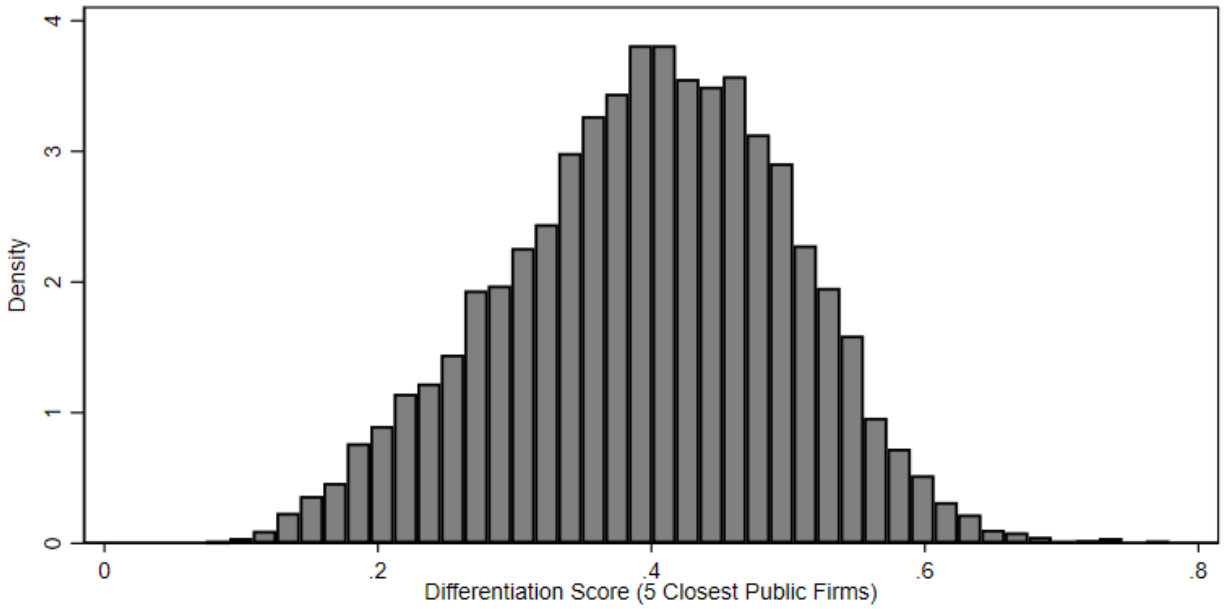
OLS model. Dependent variable is a binary variable equal to 1 if a firm is IPO or acquired and zero otherwise. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 12: Dynamic effects of founding differentiation on equity performance across age.

	(1) *Dep. Var.* IPO or Acq. (Cumulative)	(2) *Dep. Var.* IPO (Cumulative)	(3) *Dep. Var.* Acquisition (Cumulative)
Differentiation Score (5 Closest Public Firms)			
Age=0 × Differentiation Score (5 Closest Public Firms)	-0.172**	-0.00734	-0.165**
	(0.0295)	(0.00729)	(0.0275)
Age=1 × Differentiation Score (5 Closest Public Firms)	-0.154**	-0.00572	-0.148**
	(0.0282)	(0.00739)	(0.0266)
Age=2 × Differentiation Score (5 Closest Public Firms)	-0.109**	-0.00255	-0.107**
	(0.0252)	(0.00756)	(0.0236)
Age=3 × Differentiation Score (5 Closest Public Firms)	-0.0523**	0.00201	-0.0543**
	(0.0235)	(0.00777)	(0.0217)
Age=4 × Differentiation Score (5 Closest Public Firms)	0.00205	0.00663	-0.00458
	(0.0221)	(0.00829)	(0.0205)
Age=5 × Differentiation Score (5 Closest Public Firms)	0.0503**	0.00917	0.0411*
	(0.0203)	(0.00842)	(0.0197)
Age=6 × Differentiation Score (5 Closest Public Firms)	0.0900**	0.0141	0.0759**
	(0.0237)	(0.00833)	(0.0230)
Age=7 × Differentiation Score (5 Closest Public Firms)	0.121**	0.0182*	0.103**
	(0.0288)	(0.00904)	(0.0271)
Age=8 × Differentiation Score (5 Closest Public Firms)	0.144**	0.0210**	0.123**
	(0.0340)	(0.00957)	(0.0310)
Age=9 × Differentiation Score (5 Closest Public Firms)	0.163**	0.0232**	0.139**
	(0.0391)	(0.00999)	(0.0352)
Age=10 × Differentiation Score (5 Closest Public Firms)	0.175**	0.0252**	0.149**
	(0.0428)	(0.0104)	(0.0382)
Year F.E.	Yes	Yes	Yes
HP Industry F.E.	Yes	Yes	Yes
Observations	153813	153813	153813
R^2	0.073	0.059	0.071

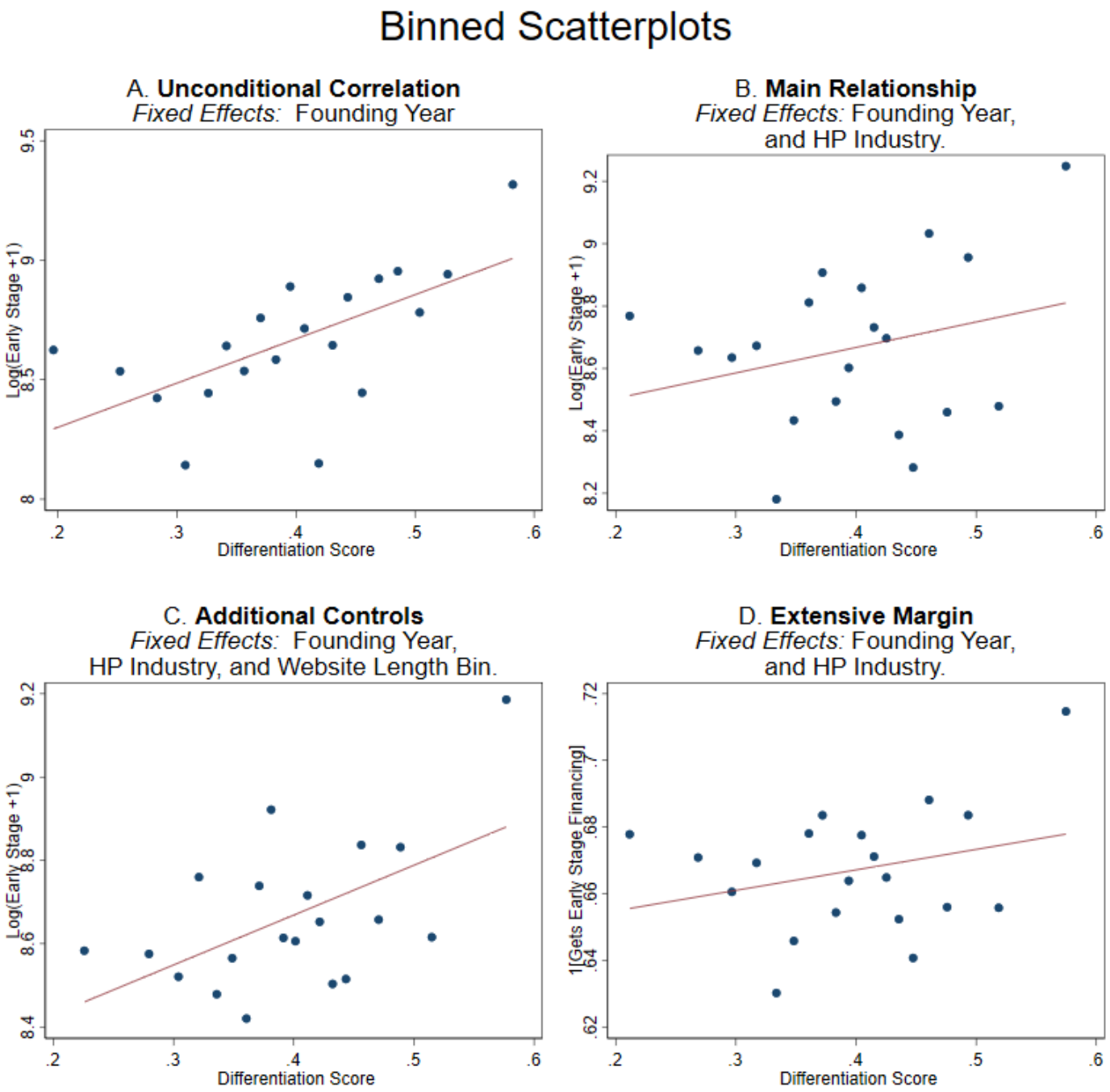
OLS model. Dependent variable is a binary variable equal to 1 if a firm has achieved IPO or acquired by age *t* and zero otherwise. Standard errors double clustered by Hoberg-Phillips industry and state. Significant reported as: * p <0.10, ** p <0.05, *** p <0.01.

Figure 1: Distribution of Strategic Differentiation Score (5 Closest Public Firms)



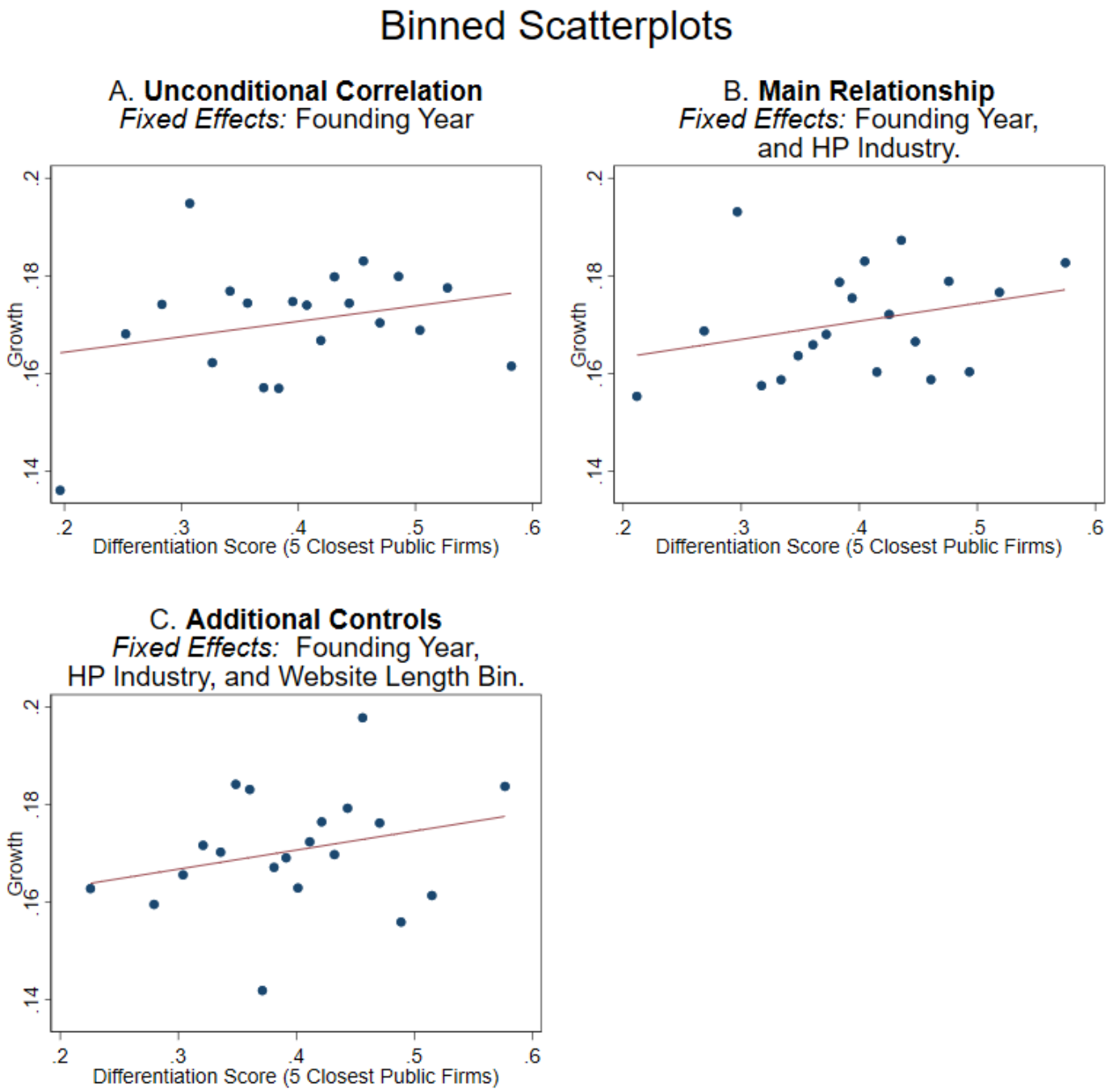
Notes: Reports the histogram of strategic differentiation score estimated as the mean distance in the founding website for the five closest public firms. Distance is one minus the similarity between websites, which is estimated using a word-embeddings algorithm of all public websites and startups in each cohort.

Figure 2: Differentiation Score and Early Stage Financing



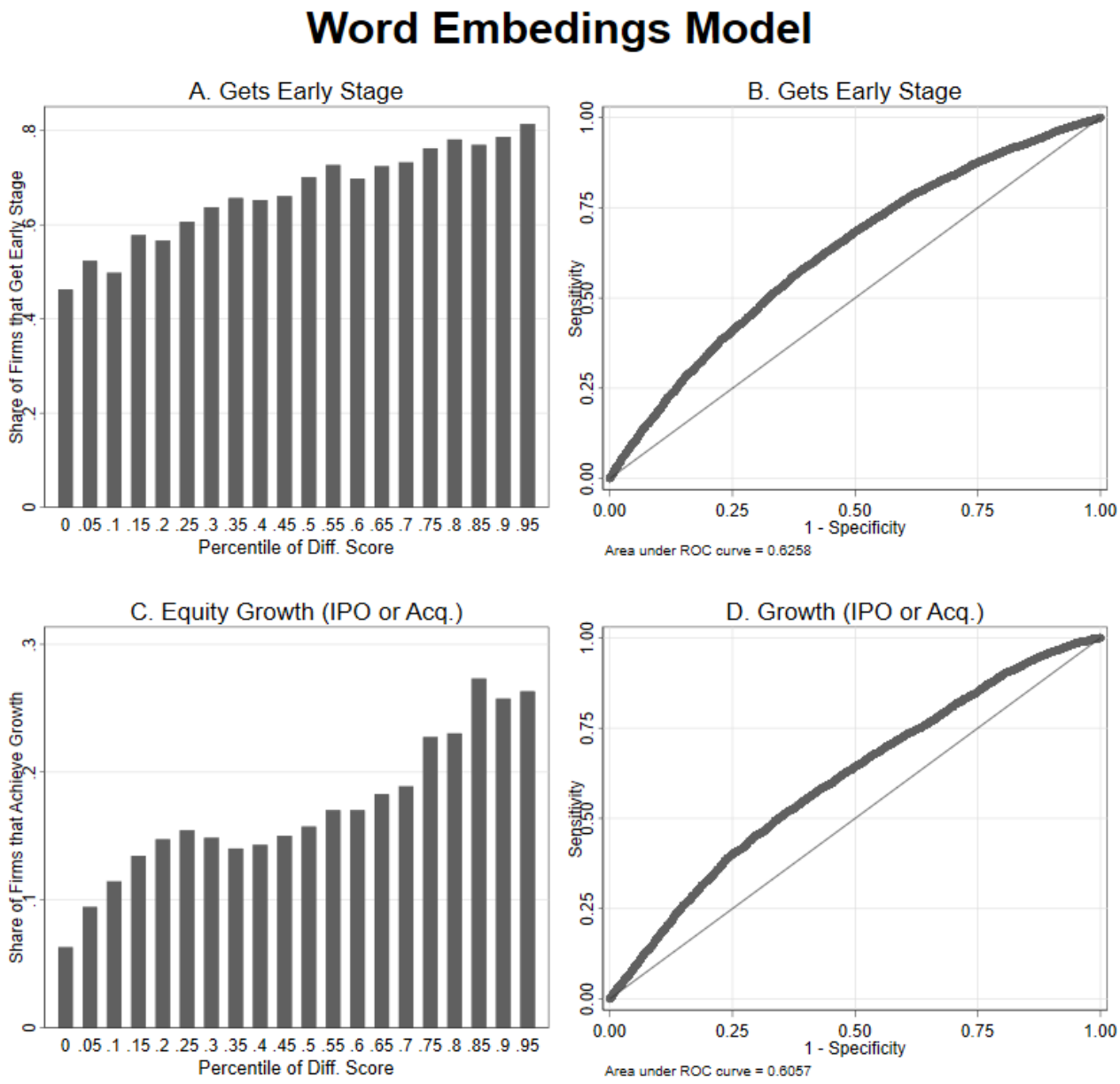
Notes: Early stage financing is all financing events recorded in Crunchbase as 'Seed', 'Angel', 'Crowdfunding', and 'Pre Seed'. HP Industry are the industries defined using the methodology of Hoberg and Phillips (2016) in our data. Appendix Figure A1 replicates these scatterplots with Series A financing events instead.

Figure 3: Differentiation Score and Equity Growth Outcomes



Notes: Growth is IPO or Acquisition. HP Industry are the industries defined using the methodology of Hoberg and Phillips (2016) in our data.

Figure 4: Out of Sample Predictability of Performance from Founding Text



Notes: This figure reports out of sample tests of how well do our measures predict performance. To do so, we run a fully interacted model of our four differentiation measures on two binary outcomes, *Gets Early Stage Financing* and *Equity Growth* using a 10-fold approach where we split the data into 10 groups and use the regression of 9 groups to predict the remaining one out of sample. Panels A and C report the distribution of outcomes across the predicted probability of performance. Panels B and D report the ROC (area under the curve) score which better measures the fit of the data.

Table A1: Examples of Strategic Differentiation Score: Food Industry

Percentile	Diff. Score (5 Closest Public Firms)	Company Name	Website	Short Description
.97	.58	www.everbowl.com	Everbowl	Everbowl is craft superfood. Everbowl specialize in create your own superfood bowls.
.94	.55	www.jicachips.com	JicaChips	JicaChips- World's 1st Jicama Chip! Less than 100 Calories Per Bag.
.88	.52	ripe.io	Ripe.io	Ripe.io is creating the Blockchain of Food.
.77	.48	www.nutpods.com	nutpods	nutpods was born out of founder Madeline Haydon's impatience for a wholesome dairy
.77	.48	www.thedropwine.com	The Drop Wine	The Drop is disrupting the wine industry by combining quality with mobile convenience and creating a Millennial-targeted lifestyle brand.
.77	.48	www.impactvi.com	ImpactVision	ImpactVision is building a new standard for food safety and quality using hyperspectral technology.
.53	.41	www.aspirefg.com	Aspire Food Group	Aspire Food Group manufactures a variety of food products made from crickets.
.51	.4	www.freshsurety.com	FreshSurety	FreshSurety enables users to wirelessly assess fresh food for spoilage without having to manually disassemble their cartons.
.49	.4	oceanshalo.com	Ocean's Halo	Ocean's Halo is a seaweed-focused food company.
.31	.35	www.betterbeanco.com	Better Bean	Better Bean offers a locally-grown line of freshly made beans in recyclable containers.
.31	.35	snowshoefood.com	Snowshoefood	SnowShoeFood develops smartphone apps that help users discover, explore and engage with food items sold through their local grocery store.
.31	.35	here.co	Here Holdings	Here Holdings creates food and beverage products in Illinois using Midwest produce.
.11	.26	www.hollison.com	Hollison Technologies	Hollison Technologies is focused on providing innovative solutions for ensuring and maintaining food safety and security.
.075	.23	www.summitwinetastings.com	Summit Wine Tastings	Founded in 2010, Summit Wine Tastings, LLC is a wine and spirits marketing and promotions company based in Chicago, Illinois.
.071	.23	oceanapproved.com	OCEAN APPROVED	OCEAN APPROVED provides domestic, fresh, healthy alternative to imported seaweed products.
.024	.18	www.bigboxoverstocks.com	Big Box Overstocks	Big Box Overstocks is a discounts store that offers discontinued, damaged packaging, end of season, oversupply, and open box items.

Table A2: Examples of Strategic Differentiation Score: Consumer Electronics

Percentile	Diff. Score (5 Closest Public Firms)	Website	Company Name	Short Description
.99	.62	parihug.com	Parihug	Parihug makes pairable, wifi-enabled teddy bears that let you hug someone from anywhere in the world.
.98	.59	www.drumpants.com	Tappur	World's first industrial quality wearable musical instrument. Watch someone play it to believe it.
.92	.54	www.joyluxinc.com	Joylux	Joylux creates innovative health solutions targeting the enormous.
.92	.53	www.lark.com	Lark	Lark is the leading digital health company using AI and clinical science to deliver scalable, positive health outcomes in chronic disease.
.89	.52	eightsleep.com	Eight Sleep	The world's first sleep fitness company.
.83	.5	www.rcski.com	RC Ski	RC Ski develops remote control technology for personal watercrafts
.83	.5	www.mistyrobotics.com	Misty Robotics	Misty Robotics, a spinoff of Sphero, is a hardware company that builds personal robots for homes and offices.
.76	.48	www.l8smartlight.com	L8 SmartLight	The L8, a device composed of 64 LED lights and a super LED light, communicates users' interests through light codes.
.72	.46	vespermems.com	Vesper	Vesper is a designer of advanced acoustic-sensing technology.
.71	.46	immediasemi.com	Immedia	Immedia Semiconductor develops and markets semiconductor based ISP and video compression technology for consumer electronics.
.55	.41	shadecraft.com	ShadeCraft	Improving human life outdoors through robotic technology .
.53	.41	asiustechnologies.com	Asius Technologies	Asius Technologies develops, protects, and markets a system of in-ear technology through its product ADEL.
.53	.41	www.werkadoo.com	Werkadoo	Werkadoo is an online portal that matches projects with candidates based on behavioral traits and soft characteristics.
.53	.41	www.ouya.tv	OUYA	OUYA develops and delivers open video game consoles for televisions.
.51	.4	ringz.tv	Ringz.TV	Ringz, an addictive video sharing application, connects users to shared playlists that are watchable on any web-connected device.
.49	.4	www.iotera.com	Iotera	Iotera was a wireless IoT networking company. It was acquired by ring.com in 2017 which was subsequently acquired by Amazon in 2018
.4	.37	www.juicequbе.com	JuiceQube	The JuiceQube is a charging station that provides the ability to charge multiple Apple, Android and other products by utilizing.
.38	.37	www.pogotec.com	PogoTec	PogoTec is the manufacturer of PogoCam. The world's smallest, lightest camera that attaches to virtually all glasses.
.29	.34	nikola.tech	Nikola	Nikola offers an advantaged far-field technology that converts radio frequency (RF) energy into usable direct current power.
.26	.33	penxy.com	Penxy	Penxy is a slide sharing application for presenters to control their presentations in real time via iOS devices.
.24	.32	hengedocks.com	Henge Docks	Henge Docks is a supplier of docking stations for Apple products.
.23	.32	swyftstore.com	Swyft	Swyft sells great products from leading brands in strategic high profile locations, serving customers.
.22	.31	www.proximaldata.com	Proximal Data	Proximal Data provides server-side caching solutions to improve server initiatives in virtualized environments.
.21	.31	neosensory.com	NeoSensory	NeoSensory gives people new senses via haptic feedback.
.21	.31	www.nextthing.co	Next Thing Co	Next Thing customizes physical things with software.
.21	.31	goplugbags.com	GoPlug	Plug Bags designs and manufactures backpacks and roller cases with a built-in high power battery.
.085	.24	www.wentworthtechnology.com	Wentworth Technology	Wentworth Technology develops SpeedThru, a drive-thru headset system for the quick service restaurant market.
.083	.24	www.unitedkeys.com	United Keys	United Keys engages in the development of technology for PC display input devices such as keyboards and keypads to private label customers.
.067	.23	www.crosswarranty.com	CrossWorld Warranty	CrossWorld Warranty is a platform for manufacturers and retail consumers to handle their warranty records online.

Figure A1: Other Measures and Early Stage Financing.

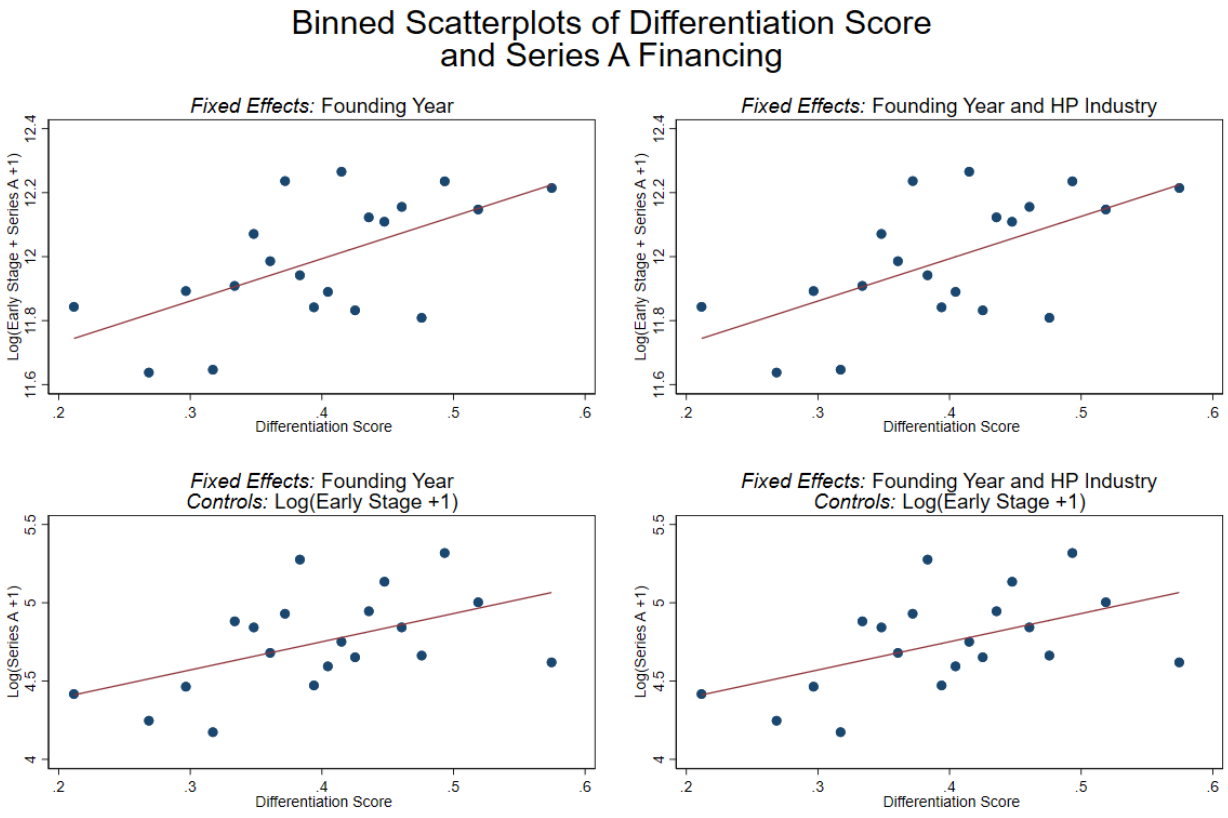


Figure A2: Other Measures and Early Stage Financing.

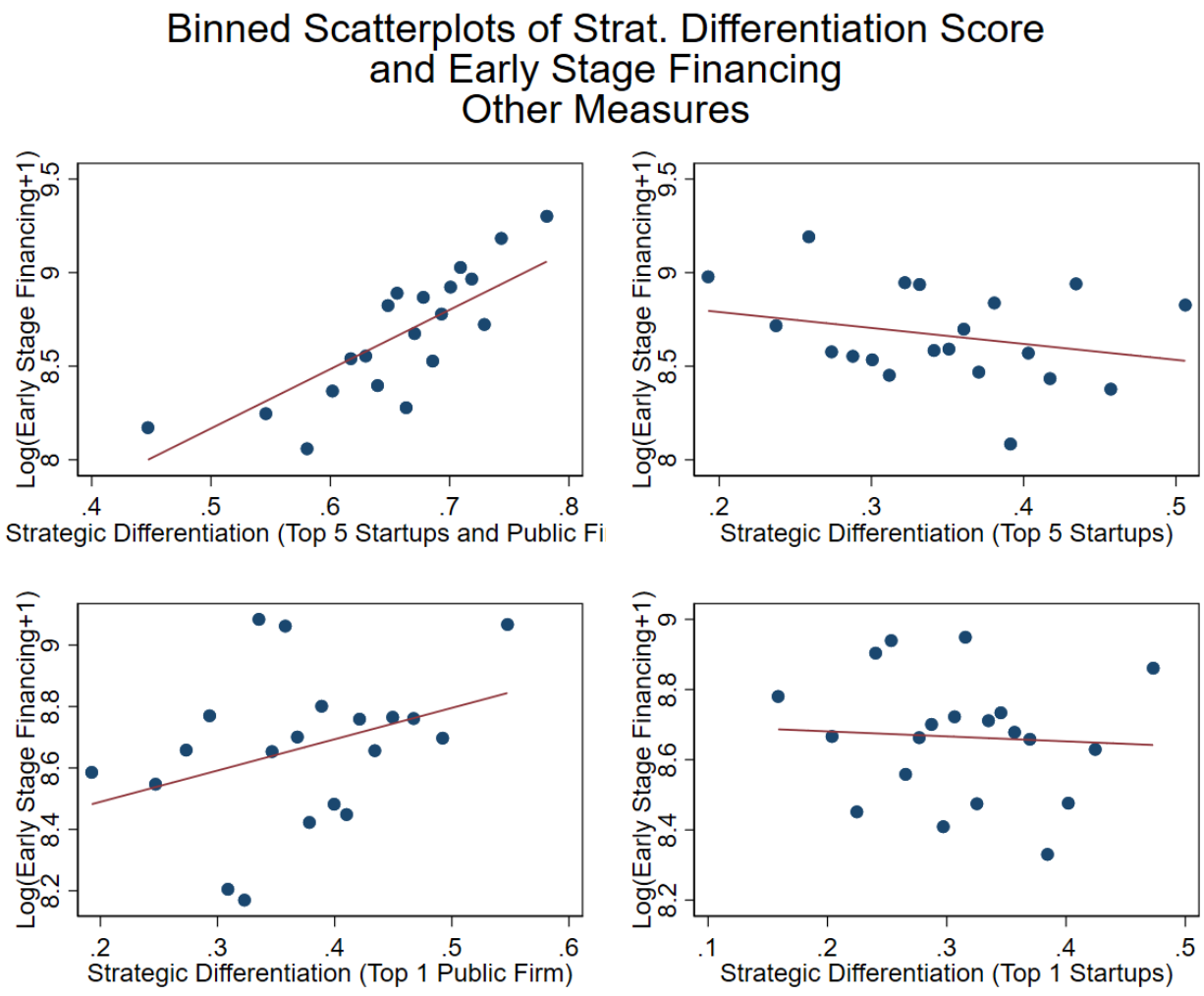


Figure A3

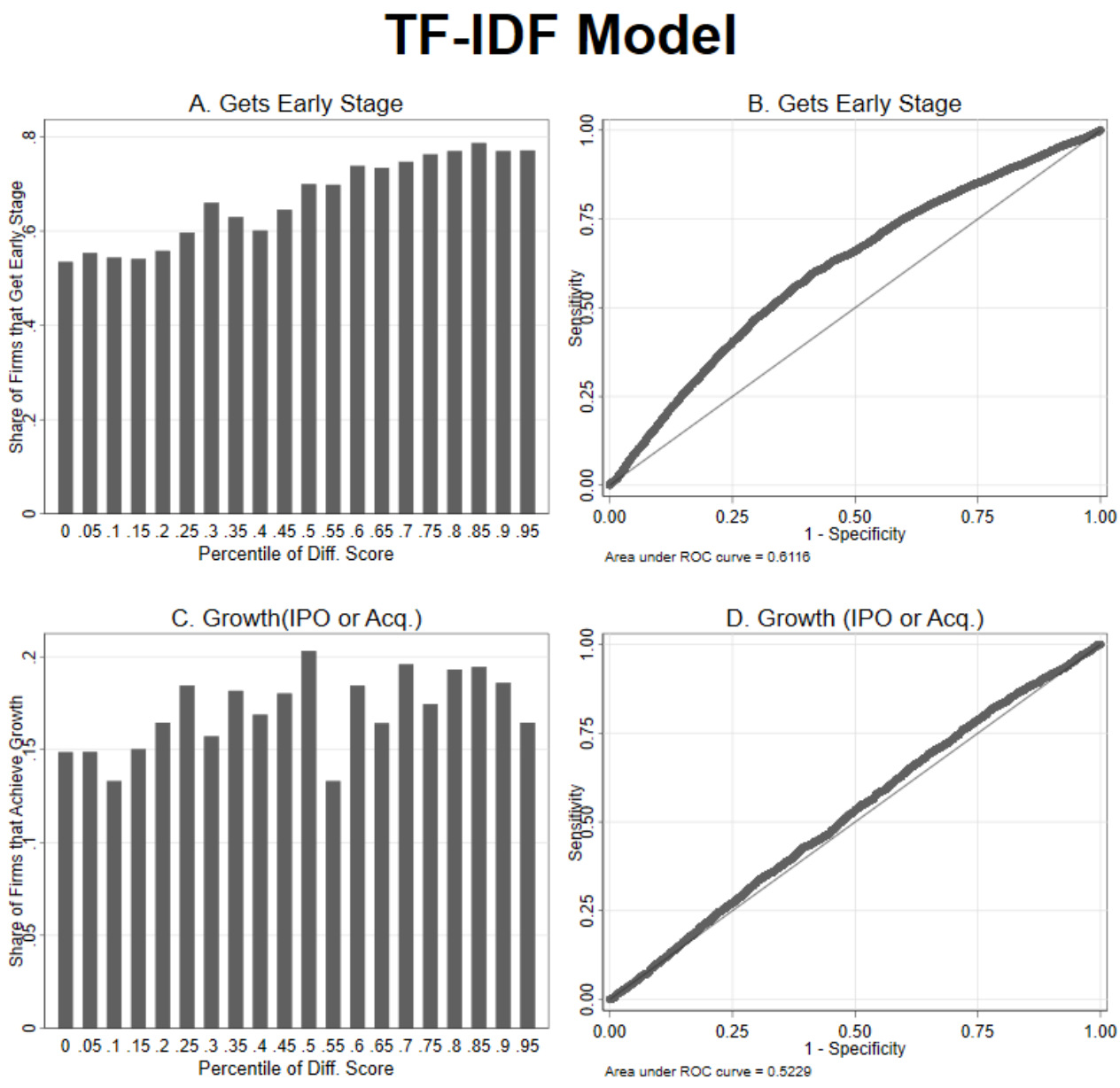
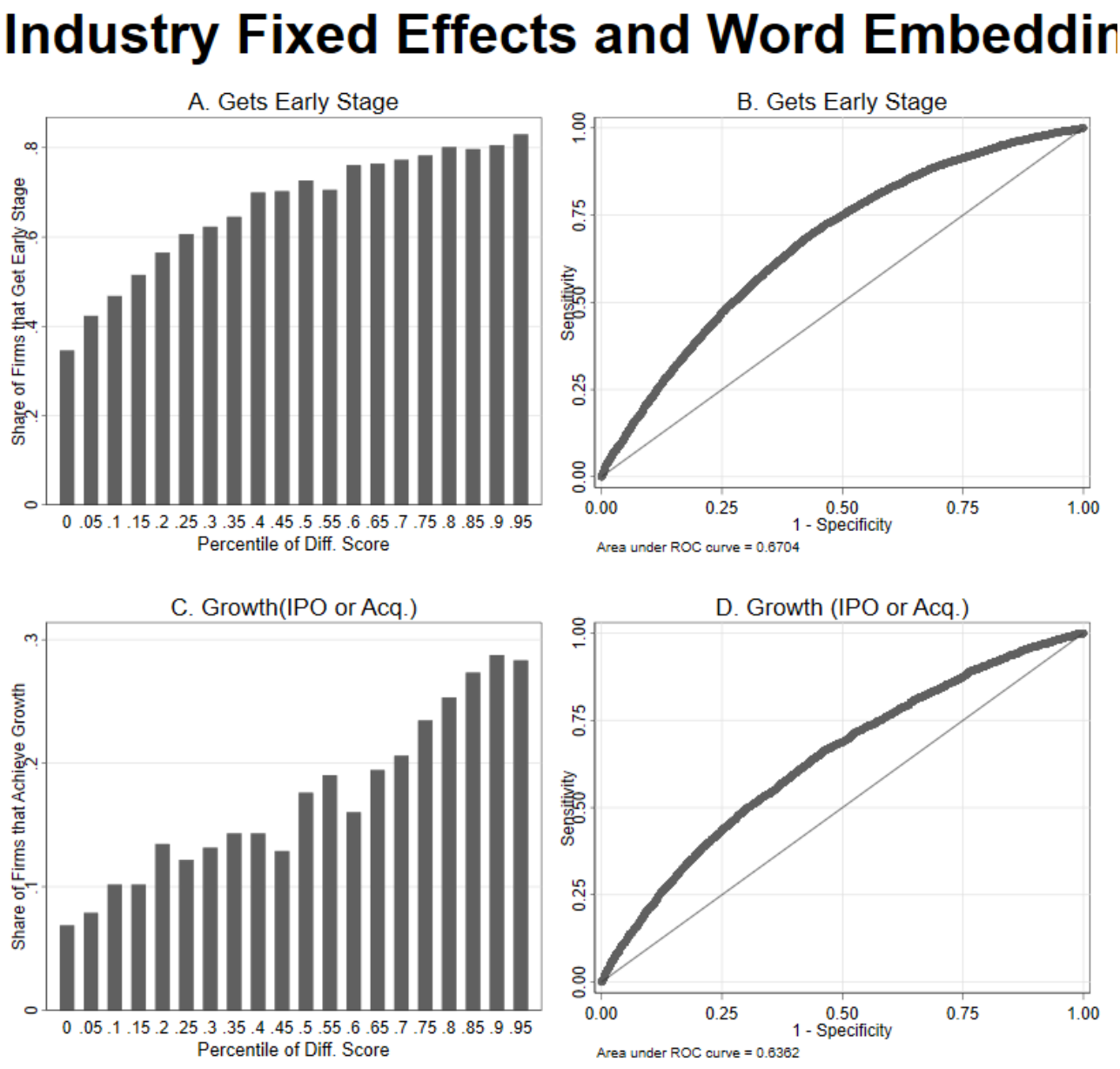

TF-IDF Model
A. Gets Early Stage
Share of Firms that Get Early Stage
0 .2 .4 .6 .8
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5 .55 .6 .65 .7 .75 .8 .85 .9 .95
Percentile of Diff. Score
B. Gets Early Stage
Sensitivity
0.00 0.25 0.50 0.75 1.00
0.00 0.25 0.50 0.75 1.00
1 - Specificity
Area under ROC curve = 0.6116
C. Growth(IPO or Acq.)
Share of Firms that Achieve Growth
0 .05 .1 .15 .2
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5 .55 .6 .65 .7 .75 .8 .85 .9 .95
Percentile of Diff. Score
D. Growth (IPO or Acq.)
Sensitivity
0.00 0.25 0.50 0.75 1.00
0.00 0.25 0.50 0.75 1.00
1 - Specificity
Area under ROC curve = 0.5229

Figure A4


Industry Fixed Effects and Word Embeddin
A. Gets Early Stage
Share of Firms that Get Early Stage
0 .2 .4 .6 .8
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5 .55 .6 .65 .7 .75 .8 .85 .9 .95
Percentile of Diff. Score
B. Gets Early Stage
Sensitivity
0.00 0.25 0.50 0.75 1.00
0.00 0.25 0.50 0.75 1.00
1 - Specificity
Area under ROC curve = 0.6704
C. Growth(IPO or Acq.)
Share of Firms that Achieve Growth
0 .1 .2 .3
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5 .55 .6 .65 .7 .75 .8 .85 .9 .95
Percentile of Diff. Score
D. Growth (IPO or Acq.)
Sensitivity
0.00 0.25 0.50 0.75 1.00
0.00 0.25 0.50 0.75 1.00
1 - Specificity
Area under ROC curve = 0.6362