With a Little Help of My (Former) Employer: Past Employment and Entrepreneurs' External Financing

Jun Huang*

Michael (Zhan) Shi†

Abstract

Securing external financing and attracting value-adding investors are challenging for nascent businesses, largely due to the concern of information asymmetry. We explore how entrepreneurs' past employment migitates this concern. Based on the rationale of inherited human and social capital, we hypothesize that: a) for startup founders, their former employers' prominence signals favorably about the quality of the new ventures, thus increasing the chances of securing funding and attracting prominent investors; b) this signaling is more relevant if the startups' business is similar to that of the former employers. Empirical tests using a novel dataset of technology companies support our hypotheses. This paper highlights the role of past employment in matching entrepreneurs with prominent investors. Methodologically, we introduce to the management literature a topic-modeling-based measure of business similarity.

*Corresponding author. Columbia Business School, 7th Floor 3022 Broadway, New York, NY 10027 USA. Email: jh3007@columbia.edu.

†W. P. Carey School of Business, Arizona State University, PO BOX 874606, Tempe, AZ 85287, USA. Email: zmshi@asu.edu.

1 Introduction

External financing is critical to the growth and success of nascent businesses. So are the value-adding services that often come along with the financial capital. In particular, with their certification and provision of guidance and resources, prominent investors are much sought after by new ventures, to the extent that entrepreneurs would offer equity at a discount in order to compete for reputable buyers.[1] The entrepreneurs' consideration can be summarized as "It is far more important whose money you get than how much you get or how much you pay for it."[2]

Despite its importance, securing external financing is a critical challenge confronting new ventures, with only a small portion getting funded and even fewer funded by prominent or reputable investors. A widely documented reason behind this challenge is information asymmetry. Investors often find it difficult to assess the quality of nascent businesses, as the latter typically lacks sufficient operating history or tangible assets. The information asymmetry causes market failure, resulting in ventures being under-financed.[3] In addition, the scarcity of prominent investors renders themselves difficult to access for many entrepreneurs.

What can entrepreneurs do to increase their chances of getting funded and to attract prominent investors? In this paper, we explore how entrepreneurs' past employers can mitigate the information asymmetry by signaling about the quality of their new ventures. As a part of the due diligence for deciding whether to fund a startup, it is common for investors to investigate the entrepreneurs' employment history. Specifically, our research questions are: 1) Do entrepreneurs that used to work for a prominent company have an edge in securing external financing, and are they more likely to attract prominent investors? 2) If former employers' prominence is indeed a favorable signal to external investors, how would that signal vary with the relation between the new ventures and the "parents" (the entrepreneurs' former employers)?

Our theoretical reasoning draws on the literature of inter-organizational endorsement. By working at a prominent company in the past, an entrepreneur inherit human and social capital that may

[1] Hsu (2004).
[2] Bygrave and Timmons (1992).
[3] Stiglitz and Weiss (1981).

be useful for her new venture. Therefore, we hypothesize that coming from a prominent former employer signals better quality of the venture, and thus are more likely to secure financing and attract prominent investors. Further, the inherited human and social capital are more relevant for the new venture if it shares a similar business model, similar practices or similar needs for resources with the parents. Therefore, we also hypothesize that the signaling by past employment to be more effective if such similarity increases.

We test the hypotheses using a novel dataset from CrunchBase, an online directory of companies, individuals and investors in the technology sector. Our sample includes both startups that are funded and those that are not. We perform coarsened exact matching to mitigate the potential endogeneity issue. As our analyses show, startups founded by former employees at more prominent companies are more likely to secure external financing than those with non-prominant parents.[4] Conditional on receiving funding, startups with prominent parents also are more likely to attract prominent investors at the initial round.[5] More detailed analyses suggest that the likelihood of securing funding declines in the order of the following groups: (1) startups with prominent parents whose business is similar to theirs, (2) startups with prominent parents that operate less similar business, (3) startups with non-prominent parents, and (4) startups whose founders had no employment experience at technology companies. Conditional on receiving funding, startups in Group (1) were also more favored by prominent investors than those in all the other groups. Overall, our results support the positive effects of parents' prominence. By showing that the effects are more pronounced for greater similarity between the startups and the parents, the findings lend support to the mechanism of entrepreneurs' inheritance from former employers.

This paper makes two contributions to the literature. First, it is the first study to show that by signaling the quality of the new ventures, entrepreneurs' past employment has implications for not only the extensive margin of external financing, but also the quality of the investors. As previous studies suggest, the "liability of newness" of a venture may be mitigated through signaling—either

[4]We define the prominence of a company by its employee representation at CrunchBase. See the empirical strategy below for details.

[5]We define the prominence of an investor by its track record of investment. See the empirical strategy below for details.

with its asset such as patents (Hsu and Ziedonis, 2013; Stinchcombe, 1965), or through affiliation with reputable exchange partners (e.g. Ozmel et al. (2012); Stuart et al. (1999)). However, we know relatively little about the effectiveness of signaling in attracting prominent investors. This paper fills this gap.

Second, in studying entrepreneurs' signaling to external investors, we contribute to the literature by showing the contingency effects of business similarity between the new ventures and the parents. Besides reaffirming the mechanism of inherited human and social capital, the contingency effects also add a boundary condition to the signaling. Methodologically, we also introduce to the management literature a measure for business similarity. The measure is based on companies' textual descriptions that summarize the nature of their business, and is generated using a topic modeling technique (Shi et al., 2014).

The rest of this paper proceeds as follows. In section, we develope the hypotheses. Section 3 introduces our empirical context and strategy. Section 4 presents the main analysis, focusing on the main effects of former employers' prominence. Section 5 investigates the contingency effects of business similarity. Finally, Section 6 concludes.

2 Theory and Hypotheses

Whereas information asymmetry is prevalent in external financing, investors find it more difficult to assess the quality of nascent ventures than that of their more mature counterparts. More mature businesses generally have a well-defined customer segment, a lineup of product offerings, a network of exchange partners and a track record of performance, all of which form the basis for evaluation of business quality. In contrast, nascent ventures typically lack some or even all of those. Funding them is thus particularly risky for external investors, who in turn demand higher compensation (e.g. discounted pricing of the equity) for taking the risks. If the ventures cannot afford to provide the demanded compensation, market failures occur, leaving the new ventures under-financed.

It is therefore important for nascent ventures to mitigate the information asymmetry to secure financing. As the examination of the "business" aspects tend to yield limited information, investors often turn to a venture's "people" aspects in assessing its value. A relatively informative and accessible aspect about the entrepreneurs is their employment history. In particular, a career stint at a prominent employer signals favorably to external investors. This signaling is grounded by the rationale that entrepreneurs that come from working for a prominent company inherit human and social capital that may be valuable for the new ventures. In effect, having a prominent employer from the past endorses the quality of the new ventures. The following elaborates the rationale.

From prominent employers, would-be entrepreneurs learn human capital that is useful for managing a new venture. Prominent companies tend to build on time-tested business practices that differentiate themselves from others. Many of those practices would help to bring a new venture on track. For example, employees may absorb commercial, technological, regulatory know-hows or certain routines when they leave to start their own ventures (Agarwal et al., 2004; Chatterji, 2009; Phillips, 2002). Those know-hows become part of the entrepreneurs' human capital, and contributes to the new ventures' success (Colombo and Grilli, 2005; Dick et al., 2013).

From prominent employers, would-be entrepreneurs also accumulate social capital that helps with the business operation of the new ventures. For example, prominent employers boast higher-caliber talent, making it easier for would-be entrepreneurs to find start-up partners (Campbell et al., 2012; Agarwal et al., 2013). Leaving as a team to start a new venture is particularly relevant when the knowledge is complicated, or is embedded in a group rather than in individuals (Ganco, 2013; Kogut and Zander, 1992).

In essence, the inheritance of human and social capital represents knowledge transfer from the parents to the new ventures. The knowledge transfer signals better quality for the new ventures. Indeed, prior studies have documented longer survival of startups where their founders used to work at incumbent firms (Eriksson and Moritz Kuhn, 2006). Therefore, the prominence of former employers can mitigate the information asymmetry that concerns external investors:

Hypothesis 1: Entrepreneurs from prominent employers are more likely to secure external

financing, compared with those from non-prominent employers or those without relevant industry experience.

Relatedly, the favorable signaling by employers' prominence implies greater attraction for prominent investors. External investors' prominence builds from their ability and experience in identifying and mentoring promising ventures (Hellmann and Puri, 2002). Given the benefits of certification and value-adding services that come along with their capital, prominent investors are scarce resources that entrepreneurs compete for. For new ventures without prior external financing, the signaling by employers' prominence implies an advantage in attracting prominent investors. Therefore:

Hypothesis 2: Conditional on receiving external financing, entrepreneurs from prominent employers are more likely to be funded by prominent investors at the initial round, compared with those from non-prominent employers or those without relevant industry experience.

The inherited human and social capital are more relevant to the new venture if the nature of its business is more similar to that of the entrepreneurs' former employers. On the one hand, it is possible that entrepreneurs inherit knowledge that may be broadly applied to many industries. On the other, some knowledge are specific to business ideas, models or practices, and would not be useful for the entrepreneurs if their new ventures have little overlap with the former employers. Therefore:

Hypothesis 3: The effects in Hypotheses 1 and 2 are more pronounced if the entrepreneurs' new ventures are more similar to their former employers in the nature of the business.

3 Empirical Strategy

3.1 *Empirical Context*

We test the hypotheses using data collected from CrunchBase in July 2013. CrunchBase is an online directory of companies, individuals and investors in the technology sector. Its spectrum of coverage ranges from the "hard-tech" industries such as hardware and semiconductor to the "soft-tech" ones such as mobile applications. The database updates its content in two ways. First, it automatically gathers information from a list of popular websites that reports the dynamics of the technology sector. Second, as a free and open platform, it allows registered users to edit its content. Before becoming public, all edits are subject to approval by the administrators. Since its launch, CrunchBase has gained increasing popularity in the technology startup community, and has become a major source of information for both entrepreneurs and investors.

The more commonly-used startup databases, such as Thompson One (formerly known as "VentureXpert") and VentureSource, typically cover only companies that are backed by venture capital (VC) firms. In contrast, CrunchBase has the distinct advantage of covering both companies that are externally funded and those that are not. As externally funded startups account for only a small portion of the population, and are probably sampled from the right tail of the quality distribution, using CrunchBase provides a more representatitive picture of the technology startups. In addition, specific to the purpose of this study, the commonly-used databases lack variation on the extensive margin of external financing.

The coverage of CrunchBase is comparable to that of the other data sources. For instance, it covers about the same number of investment deals in the corresponding sector as recorded by the National Venture Capital Association (Block and Sandner, 2009).

In this study, we restrict the sample to startups founded between the start of 2007 and the end of 2012. As CrunchBase was launched in 2007, the content before 2007 was added retrospectively and thus may not be complete; As the data were collected in mid-2013, we limit our analysis to startups founded by the end of the previous year because the very nascent entries are likely to bias our estimators towards 0 (Shi et al., 2014). A detailed overview of the sample is presented with the analyeses below.

3.2 *Key Measures*

A key aspect in testing the hypotheses is to operationalize the "prominence" of the companies and of the investors. For the companies, we measure their prominence by the total number of employees (past or present) that are registered at CrunchBase. This prominence measure captures a company's representation and influence in the technology startup community. Out of the total 155,000 technology companies featured at CrunchBase, we define "stars" to be the 100 with the most employees registered at CrunchBase. Table 1 presents the top and bottom 20 of the 100 stars. The list of stars features mostly household names in the technology startup community, lending credence to our measure of prominence. We further define "star-spawns" as startups with at least one founder being a former employee at a star company.[6]

For the investors (both institutions and individuals), we measure their prominence by the total number of exits since the start of 2007. Exits include both merger and acquisition and initial public offering. This prominence measure is based on an investor's performance, which reflects its status in the technology startup community. Out of the total 2,351 investors with at least one exit, we define prominent investors to be the top 10 percent of them by the number of exits.[7] Table 2 presents the top and bottom 20 of the prominent investors. The list of names is consistent with the common perception of reputable investors.[8]

[6] One may also specify a continuous measure for the prominence of the founders' employers. As this paper employs matching (presented in the empirical strategy below), we use a binary measure for this variable.

[7] The cut-off threshold is 5 exits.

[8] Alternatively, we measure prominence by an investor's number of investments or number of companies invested. These measures are highly correlated with the one we presented in the main text. The correlation coefficients are above 0.9.

3.3 *Analyses*

We conduct two analyeses. In the main analysis, we test Hypotheses 1 and 2 by estimating the effects of employment prominence on the likelihood of securing external financing and on the prominence of the investors. In the additional analysis, we test Hypothesis 3 through breaking down the star-spawns by their business similarity with their founders' former employers. For each analysis, we perform coarsened exact matching to mitigate the concern for endogeneity.

4 Main Analysis: Star-Spawns vs. Other Spawns

4.1 *Matched Sample*

For a startup, being a star-spawn is associated with multiple factors that cannot be observed but potentially affects its external finanicng. These factors may include unobserved founders' talent (e.g. smarter people work for more prominent companies), and unobserved quality of the business idea (i.e. as the opportunity cost may be higher for leaving a prominent company, then only those with really good ideas would quit). In statistical analysis, these factors can create endogeneity issues that bias the estimators upwards. To estimate the causal effect of employment prominence, ideally we would like to randomly assign prominent employers to individuals and observe their subsequent entrepreneurial behaviors. Unfortunately it seems difficult to do so, because we are not aware of any reputable company that accepts randomly assigned intakes, and also because we cannot force company employees into entrepreneurship.

Instead of pursuing an experiment, we employ coarsened exact matching (CEM) to mitigate the potential endogeneity. In the main analysis, we compare two groups of startups: star-spawns and the others. CEM works by first producing a number of strata based on the discrete values of the specified observables. It keeps only the strata that contains startups from both groups, and drops the other strata.[9] In this way, CEM enhances the balance between the two groups. Compared with

[9] Within each stratum, we allow 1 to N matching.

the original sample, the more balanced sample would theoretically weaken the association between the confounding factors and the employment prominence, thus mitigating endogeneity.

We match the star-spawns and the other startups based on the following observables: the number of founders, whether any founder had past entrepreneurship experience, whether any founder used to work at the technology company, the startup's geographical location, its founding year and industry. The geographical location is denoted by state for U.S.-based startups, and denoted by country for startups outside the U.S. In addition, at the time of our collection, the CrunchBase database no longer applies industry labels to the companies. We derive the industry classification from the startups' description. For most startups, a piece of text (typically in a couple of paragraphs) describes the nature of its business. On these texts, we run a topic modeling technique termed Latent Dirichlet Allocation (LDA) (Blei et al., 2003). The startups are then each classified into one of 25 industries.[10]

The matched sample contains 3,165 observations. 769 of them are star-spawns. The rest are the other spawns, that is, startups where no founder used to work at a star company but at least one was employed at a technology company before. Table 3 presents an overview of the matched sample. Star-spawns were more likely to secure external financing than the other spawns. The difference is 51 vs. 36 percentage points. Conditional on receiving funding, star-spawns were also more likely to be funded by prominent investors at the initial round. The difference is 62 vs. 46 percentage points. On average, both star-spawns and the other spawns had about 2 founders. Around 40 percent of the startups were founded by serial entrepreneurs. A higher portion of star-spawns were based in the U.S. and in California than the other spawns, but they both share similar distributions over the founding years. Figure 1 clearly shows that at any time point since birth, star-spawns were more likely to secure funding.

[10] See the Appendix for the list of the industries and an introduction to LDA.

4.2 *Regressions*

All the regressions in this paper are weighted by the size of the strata. To account for right censoring, we use Cox proportional-hazards model to estimate the effect of employment prominence on the extensive margin of external financing. In the model, the event is receiving the first external financing. The hazard function takes the following form:

$$\lambda_i(t) = \lambda_i^0(t) \exp(\alpha + StarSpawn_i \cdot \beta + \mathbf{X_i} \cdot \gamma)$$

where $\lambda_i^0(t)$ denotes the baseline hazards for startup i at time t. $StarSpawn_i$ is a binary indicator for being a star-spawn. $\mathbf{X_i}$ is a set of control variables including the number of founders, whether any founder had prior entrepreneurial experience and fixed effects for industry, geography and founded year. Among the coefficients α, β and γ, we are primarily interested in β, which denotes the effect of employment prominence.

Conditional on receiving funding, we use logistic regression to study the effect on the quality of the first-round investors. We focus on the initial round because the subsequent rounds (if any) are likely to be affected by the investors in the previous rounds. The model is as follows:

$$logodds(y_i) = \alpha + StarSpawn_i \cdot \beta + \mathbf{X_i} \cdot \gamma$$

where y_i denotes the event that startup i's first-round investors include at least one prominent investor. The other notations are the same as above.

The regression results are summarized in Table 4. In both the Cox and the logistic regressions, the coefficient for star-spawn is positive and statistically significant. The results suggest that compared with other spawns, star-spawns were more favored by external investors, and had a higher chance of attracting prominent investors at the initial round. The statistical findings support Hypotheses 1 and 2.

5 Additional Analysis: All Startups

5.1 *Breakdown of All Startups*

For a more comprehensive understanding of the effects of entrepreneurs' past employment, we compare the star-spawns with not only other spawns, but also non-spawns, *i.e.*, startups whose founders had no prior experience at technology companies. To test Hypothesis 3, we break down star-spawns by their business similarity with their founders' former employers.

To measure business similarity between two companies, we compute cosine similarity using the weights of the industries. The weights are produced by Latent Dirichlet Allocation using the companies' description.[11] Thus, the dyadic business similarity measure is bounded between 0 and 1, with larger values indicating greater similarity in terms of the nature of the business. In this analysis, we categorize the star-spawns into "high-similarity" ones, *i.e.* those that are more similar to their founders' former employers (business similarity $\geq$0.5) and "low-similarity" ones, *i.e.*, those that are less so (business similarity $<$0.5).

For the four categories including high-similarity star-spawns, low-similarity star-spawns, other spawns and non-spawns, we perform coarsened exact matching based on the following observables: the number of founders, the startup's geographical location, its founding year and industry. The matched sample contains 6,522 observations, of which 73 are high-similarity star-spawns, 360 are low-similarity star-spawns, 1332 are other spawns and the rest are non-spawns. Table 5 presents an overview of the sample. The descriptive statistics share a similar pattern to those in the main analysis. Notably, among the four groups of startups, high-similarity star-spawns were the most likely to secure external financing. Conditional on receiving funding, they are also more likely to attract prominent investors at the initial round.

Figure 2 shows the group comparison by the startups' age. In terms of likelihood of securing external funding, spawns outperform non-spawns, star-spawns outperform the other spawns, and high-similarity star-spawns outperform their low-similarity counterparts. This ranking emerges

[11] Please refer to the appendix for the technical details.

since the startups' very nascent stage, and persists over time.

5.2 *Regressions*

As in the main analysis, we apply Cox proportional-hazards model to study the effect on extensive margin of financing. The model is:

$$\lambda_i(t) = \lambda_i^0(t)\exp(\alpha + HighStarSpawn_i \cdot \beta_1 + LowStarSpawn_i \cdot \beta_2 + OtherSpawn_i \cdot \beta_3 + \mathbf{X_i} \cdot \gamma)$$

where $HighStarSpawn_i$, $LowStarSpawn_i$, $OtherSpawn_i$ are binary indicators. The other variables are as defined above. In this model, the baseline group is the non-spawns. We are primarily interested in the coefficients β_1, β_2, and β_3, which represent the comparison with the baseline group. We are also interested in comparing β_1 and β_2. This comparison informs how the effects of former employers' prominence vary by business similarity with the parents.

To study the effect on the quality of first-round investors, we apply the following logistic regression model:

$$logodds(y_i) = \alpha + HighStarSpawn_i \cdot \beta_1 + LowStarSpawn_i \cdot \beta_2 + OtherSpawn_i \cdot \beta_3 + \mathbf{X_i} \cdot \gamma$$

where the variables are as defined above.

The regression results, as summarized in Table 6, support Hypothesis 3. The estimates for the β's are positive, and their magnitude supports the ranking in the descriptive statistics. Notably, the coefficients for high-similarity star-spawns are significantly greater than those for low-similarity star-spawns and other spawns. This suggests that the effects of former employers' prominence increase with business similarity.

6 Conclusion

In sum, our statistical analyses support the hypotheses. For entrepreneurs, having worked for a prominent employer before signals favorably to external investors. This helps them to raise early-stage funding and attract high-quality investors. We also find that the effects are more pronounced if the new ventures are in a similar business with the former employers. This lends further support to our rationale that entrepreneurs inherit human and social capital from their former employers.

The findings have several important implications. First, they deepen our understanding of the value of human assets (Coff, 1997). The founders' and employees' human and social capital are a major source of nascent ventures' competitive advantage (Klepper and Sleeper, 2005; Agarwal et al., 2007; Acs et al., 2009). We provide direct evidence that the capital market recognizes the value of inter-organizational knowledge spillover in the context of entrepreneurial firms. Second, this paper also has implications for individuals' professional career choice. As the empirical results suggest, on average, the external investors value entrepreneurs' prior employment experience, and also value the "brand name" of the prior employers. Potential entrepreneurs may find these results helpful when considering whether to work for a company before startup. Further, the results also shed light on the question of "who to work for before startup". An array of former employers' characteristics have been documented to influence entrepreneurs' competence. For example, some find that working for smaller employers facilitates the cultivation of entrepreneurial skills (Elfenbein et al., 2010; Sørensen and Phillips, 2011), others research has found little size effect (Dick et al., 2013). This paper adds a related but somewhat different perspective: compared with working for other companies, employment experience at prominent companies confer an advantage in raising early-stage capital.

Appendix: Industry Classification and Business Similarity

In this paper, we derive the industry classification and compute dyadic business similarity using the startups' description at CrunchBase. The description is typically a piece of text, usually in a couple of paragraphs, that summarizes the business of the company. As an illustrating example, the following is the description for the company Twitter:

```
   Twitter is a global social networking platform that allows its users
to send and read 140-character messages known as ‘‘tweets’’.  It enables
registered users to read and post their tweets through the web, short
message service (SMS), and mobile applications.
   As a global real-time communications platform, Twitter has more than
400 million monthly visitors and 255 million monthly active users around
the world.  Twitter’s active group of registered members includes World
leaders, major athletes, star performers, news organizations, and entertainment
outlets.  It is currently available in more 35 languages.
   Twitter was launched in 2006 by Jack Dorsey, Evan Williams, Biz Stone,
and Noah Glass.  It is headquartered in San Francisco, C.A. with local
offices in Atlanta, Austin, Boston, Boulder, Chicago, Detroit, Los Angeles,
New York, Seattle, Sunnyvale, and Washington.  Twitter’s international
offices are located in Amsterdam, Berlin, Dublin, London, Madrid, Paris,
Rio de Janeiro, São Paulo, Singapore, Sydney, Seoul, Tokyo, Toronto,
and Vancouver.
```

For an individual company, we first use Latent Dirichlet Allocation (LDA), a topic modeling technique, to generate a series of weights over a specified number of industries. The weights are then used to determine the primary industry for the company, as well as to produce a cosine similarity measure for a pair of companies. LDA falls into the category of unsupervised machine learning, and has been used for classification in the computer science literature(Fang et al., 2013). Recently, LDA has also started to be used in the economics literature (Hansen et al., 2014). It works as follows:

A description d is a set of words $\{w_d^j | j = 1, 2, ..., L_d\}$, where L_d is the number of words in d. Let D be the number of descriptions in the sample. Let W be the set of all the words in all the descriptions. We specify K industries. Industry $k \in \{1, 2, ..., K\}$ is characterized by p_k, a probabilistic distribution over W. We denote p_k^w as the probability of the word w appearing in the description if the industry is k. The likelihood of the description d belonging to a company in industry k is q_d^k. Then, the likelihood of d being observed is a product of the likelihood of its individual words being observed:

$$\prod_{j=1}^{L_d}(\sum_{k=1}^{K} q_d^k \cdot p_k^{w_d^j})$$

For computational efficiency, LDA estimates p and q using Bayesian estimation. Assuming Dirichlet priors for p and q with parameters θ and δ respectively. Then, the joint likelihood function is:

$$\prod_{k=1}^{K} Pr(p_k|\theta) \cdot \prod_{d=1}^{D} Pr(q_d|\delta) \cdot \prod_{j=1}^{L_d}(\sum_{k=1}^{K} q_d^k \cdot p_k^{w_d^j})$$

From that we may derive the posterior distribution using Markov Chain Monte Carlo. The estimates produce a weight for each word in each industry. For a given company, from its description and the estimated weights on words, we can generate a weight for each industry. The industry with the highest weight is designated as the company's primary industry.

In this paper, we let $K = 25$.[12] To (roughly) profile each industry, Table 7 lists the five words with the greatest weights. It also shows how the industries are distributed in the sample. At CrunchBase, the most popular industry seems to be related to fashion retailers.

From the K industry weights for each company, we can compute the dyadic business similarity as the cosine similarity between company i and j:

$$BusinessSimilarity_{i,j} = \frac{\sum_{k=1}^{K} W_{i,k} W_{j,k}}{\sqrt{\sum_{k=1}^{K} W_{i,k}^2} \cdot \sqrt{\sum_{k=1}^{K} W_{j,k}^2}}$$

where $W_{i,k}$ is the company i's weight in industry k. The business similarity measure is bounded between 0 and 1, with larger values indicating greater similarity between two companies' business.

[12] Alternatively, we let $K = 50$ and the statistical results are similar.

Figure 3 shows the average business similarity between star-spawns and their parents. Since 2007, we observe an interesting declining trend in business similarity. There may be multiple explanations for this observation, and it is beyond the scope of this paper to investigate the reasons behind this. One potential explanation may be the genesis of new business models (e.g. peer-to-peer lending, virtual currencies in recent years) or variation of existing business models (e.g. social network for niche populations, e-commerce for niche markets).

References

Acs, Z. J., P. Braunerhjelm, D. B. Audretsch, and B. Carlsson (2009). The knowledge spillover theory of entrepreneurship. *Small business economics 32*(1), 15–30.

Agarwal, R., D. Audretsch, and M. Sarkar (2007). The process of creative construction: knowledge spillovers, entrepreneurship, and economic growth. *Strategic Entrepreneurship Journal 1*(3-4), 263–286.

Agarwal, R., B. Campbell, A. Franco, and M. Ganco (2013). What do i take with me?: The mediating effect of spin-out team size and tenure on the founder-firm performance relationship. *US Census Bureau Center for Economic Studies Paper No. CES-WP-13-17*.

Agarwal, R., R. Echambadi, A. M. Franco, and M. Sarkar (2004). Knowledge transfer through inheritance: Spin-out generation, development, and survival. *Academy of Management Journal 47*(4), 501–522.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *the Journal of machine Learning research 3*, 993–1022.

Block, J. and P. Sandner (2009). What is the effect of the financial crisis on venture capital financing? empirical evidence from us internet start-ups. *Venture Capital 11*(4), 295–309.

Bygrave, W. D. and J. A. Timmons (1992). *Venture capital at the crossroads*. Harvard Business Press.

Campbell, B. A., M. Ganco, A. M. Franco, and R. Agarwal (2012). Who leaves, where to, and why worry? employee mobility, entrepreneurship and effects on source firm performance. *Strategic Management Journal 33*(1), 65–87.

Chatterji, A. K. (2009). Spawned with a silver spoon? entrepreneurial performance and innovation in the medical device industry. *Strategic Management Journal 30*(2), 185–206.

Coff, R. W. (1997). Human assets and management dilemmas: Coping with hazards on the road to resource-based theory. *Academy of Management Review 22*(2), 374–402.

Colombo, M. G. and L. Grilli (2005). Founders' human capital and the growth of new technology-based firms: A competence-based view. *Research policy 34*(6), 795–816.

Dick, J. M., K. Hussinger, B. Blumberg, and J. Hagedoorn (2013). Is success hereditary? evidence on the performance of spawned ventures. *Small Business Economics 40*(4), 911–931.

Elfenbein, D. W., B. H. Hamilton, and T. R. Zenger (2010). The small firm effect and the entrepreneurial spawning of scientists and engineers. *Management Science 56*(4), 659–681.

Eriksson, T. and J. Moritz Kuhn (2006). Firm spin-offs in denmark 1981–2000: patterns of entry and exit. *International Journal of Industrial Organization 24*(5), 1021–1040.

Fang, F., K. Dutta, and A. Datta (2013). Lda-based industry classification.

Ganco, M. (2013). Cutting the gordian knot: The effect of knowledge complexity on employee mobility and entrepreneurship. *Strategic Management Journal 34*(6), 666–686.

Hansen, S., M. McMahon, and A. Prat (2014, June). Transparency and Deliberation within the FOMC: A Computational Linguistics Approach. CEP Discussion Papers dp1276, Centre for Economic Performance, LSE.

Hellmann, T. and M. Puri (2002). Venture capital and the professionalization of start-up firms: Empirical evidence. *The Journal of Finance 57*(1), 169–197.

Hsu, D. H. (2004). What do entrepreneurs pay for venture capital affiliation? *The Journal of Finance 59*(4), 1805–1844.

Hsu, D. H. and R. H. Ziedonis (2013). Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents. *Strategic Management Journal 34*(7), 761–781.

Klepper, S. and S. Sleeper (2005). Entry by spinoffs. *Management science 51*(8), 1291–1306.

Kogut, B. and U. Zander (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science 3*(3), pp. 383–397.

Ozmel, U., J. Reuer, and R. Gulati (2012). Signals across multiple networks: How venture capital and alliance networks affect interorganizational collaboration. *Academy of Management Journal*, amj–2009.

Phillips, D. J. (2002). A genealogical approach to organizational life chances: The parent-progeny transfer among silicon valley law firms, 1946–1996. *Administrative Science Quarterly 47*(3), 474–506.

Shi, Z., G. M. Lee, and A. B. Whinston (2014). Towards a better measure of business proximity: Topic modeling for analyzing m&as.

Sørensen, J. B. and D. J. Phillips (2011). Competence and commitment: employer size and entrepreneurial endurance. *Industrial and Corporate Change*, dtr025.

Stiglitz, J. E. and A. Weiss (1981). Credit rationing in markets with imperfect information. *The American economic review*, 393–410.

Stinchcombe, A. L. (1965). Social structure and organizations. *Handbook of organizations*, 142–193.

Stuart, T. E., H. Hoang, and R. C. Hybels (1999). Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative science quarterly 44*(2), 315–349.

Figure 1: Likelihood of Securing Funding: Star-Spawns vs. Other Spawns

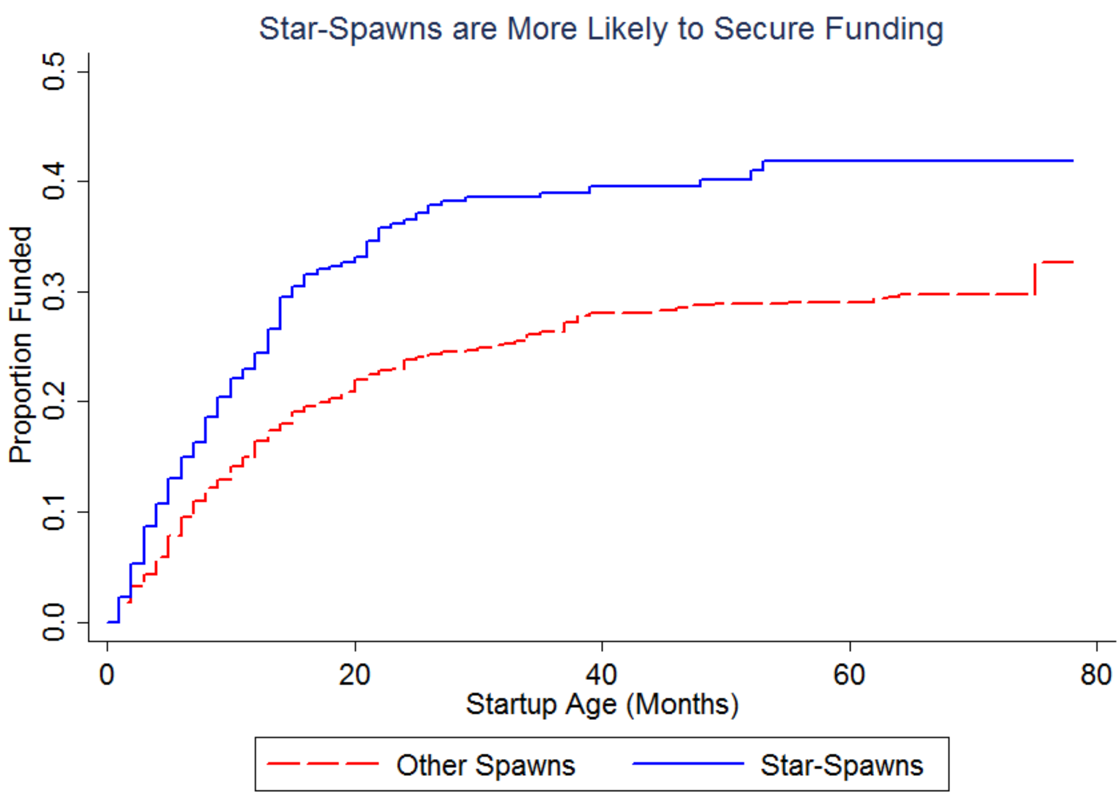

Figure 2: Likelihood of Securing Funding: All Startups

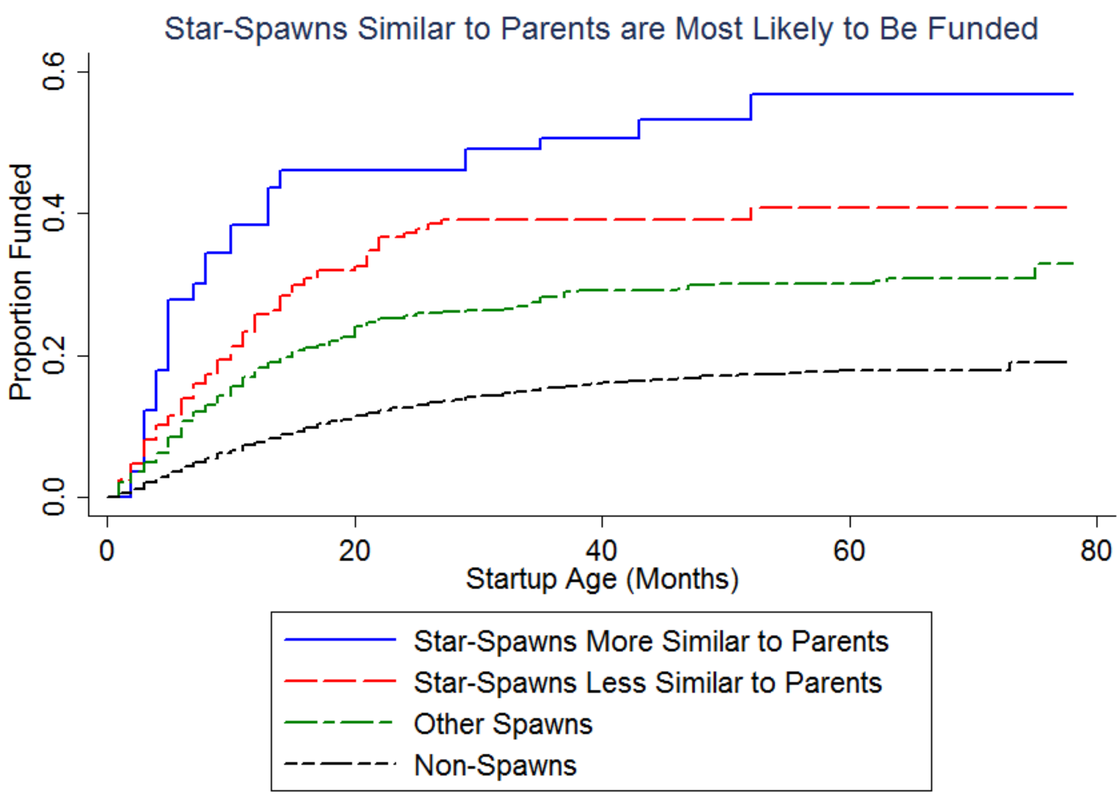

Figure 3: Average Business Similarity Between Star-Spawns and Parents

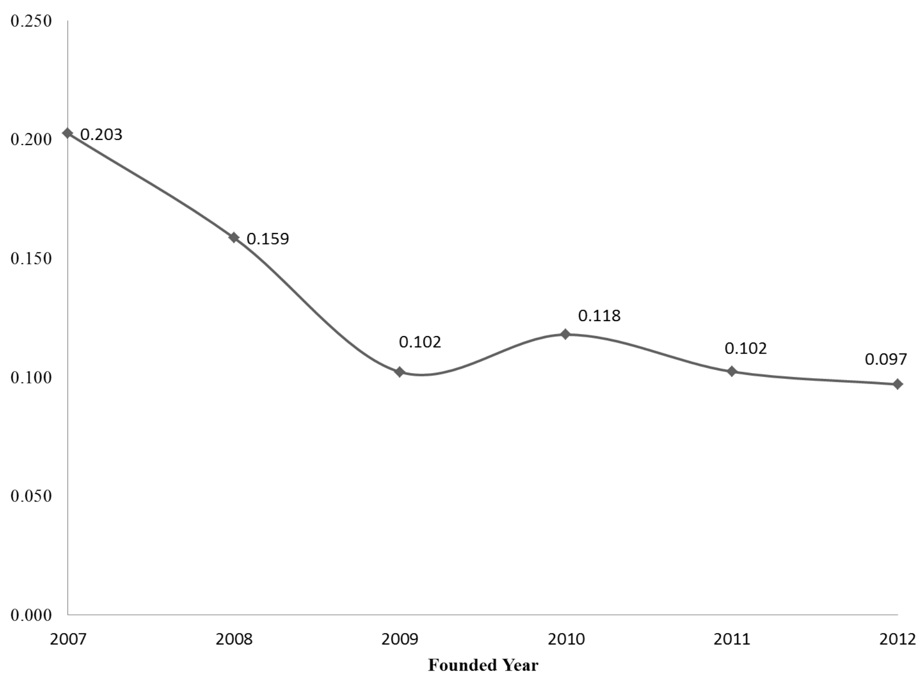

Table 1: List of Prominent Companies: Top 20 and Bottom 20

Top 20			Bottom 20		
Rank	Company Name	Representation	Rank	Company Name	Representation
1	Yahoo!	962	81	Genentech	49
2	Google	884	82	Intridea	49
3	Microsoft	842	83	McAfee	49
4	IBM	566	84	Sonicbids	49
5	Oracle Corporation	399	85	SocialMixr	49
6	Hewlett -Packard	363	86	Bloombergs	48
7	Apple	340	87	Compaq	48
8	Cisco	323	88	HubSpot	47
9	AOL	308	89	Good Technology	46
10	eBay	278	90	Demand Media	45
11	Facebook	234	91	Jive Software	44
12	Sun Microsystems	222	92	Match.com	44
13	Intel	212	93	Samsung Electronics	44
14	Amazon	187	94	6Wunderkinder	44
15	Adobe Systems	150	95	LocalResponse	43
16	EMC	144	96	betaworks	43
17	Nokia	138	97	Linden Lab	43
18	Twitter	137	98	Excite@Home	43
19	TechCrunch	129	99	CareerBuilder	43
20	PayPal	120	100	AMD	43

Table 2: List of Prominent Investors: Top 20 and Bottom 20

Top 20			Bottom 20		
Rank	Company Name	No. of Exits	Rank	Company Name	No. of Exits
1	SV Angel	73	247	InterSouth Partners	5
2	Intel Capital	46	248	Credit Suisse	5
3	Sequoia Capital	44	249	Thrive Capital	5
4	New Enterprise Associates	43	250	Cipio Partners	5
5	Accel Partners	42	251	Technology Crossover Ventures	5
6	Draper Fisher Jurvestson (DFJ)	37	252	Giza Venture Capital	5
7	Greylock Partners	34	253	Brookside Capital	5
8	First Round Capital	32	254	Rose Tech Ventures	5
9	Menlo Ventures	29	255	Sevin Rosen Funds	5
10	Charles River Ventures	28	256	Longworth Venture Partners	5
11	Kleiner Perkins Caufiled & Byers	28	257	ARCH Venture Partners	5
12	Ron Conway	25	258	Li Ka-Shing	5
13	Venrock	25	259	DFJ Gotham Ventures	5
14	500 Startups	24	260	GoldHill Capital	5
15	Redpoint Ventures	24	261	Kodiak Venture Partners	5
16	Goldman Sachs	23	262	Walden International	5
17	Ignition Partners	23	263	Tom McInerney	5
18	Benchmark	23	264	Common Angels	5
19	True Ventures	23	265	Nexus Venture Partners	5
20	Felicis Ventures	23	266	RockPort Capital Partners	5

Table 3: Descriptive Statistics: Star-Spawns vs. Other Spawns

	Star-Spawns			Other Spawns		
	mean	st. dev	N	mean	st. dev	N
Secured External Financing	0.51	0.50	769	0.36	0.48	2396
Initial Round Funded by Prominent Investors	0.62	0.49	393	0.46	0.50	871
No. of Founders	1.96	0.91	769	1.71	0.79	2396
Prior Founder Experience	0.40	0.49	769	0.45	0.50	2396
Based in U.S.A.	0.80	0.40	769	0.67	0.47	2396
Based in California	0.48	0.50	756	0.33	0.47	2362
Based in Massachusettes	0.04	0.19	756	0.04	0.18	2362
Based in New York	0.11	0.31	756	0.12	0.32	2362
Founded in 2007	0.10	0.30	769	0.10	0.30	2396
Founded in 2008	0.13	0.34	769	0.11	0.31	2396
Founded in 2009	0.17	0.37	769	0.14	0.34	2396
Founded in 2010	0.21	0.41	769	0.22	0.41	2396
Founded in 2011	0.24	0.43	769	0.26	0.44	2396
Founded in 2012	0.15	0.36	769	0.18	0.38	2396

Table 4: Regression Coefficients for Main Analysis

	Secured External Financing	Initial Round Funded by Prominent Investors
	(1) Cox Proportional-Hazards	(2) Logistic
Star-Spawn	0.406***	0.436***
	(0.099)	(0.161)
No. of Founders	0.240***	0.142
	(0.049)	(0.090)
Prior Founder Experience	0.292***	0.425***
	(0.092)	(0.156)
Industry Fixed Effects	Yes	Yes
Geography Fixed Effects	Yes	Yes
Founded Year Fixed Effects	Yes	Yes
χ^2	421.35	159.77
Log Likelihood	-3840.85	-680.90
N	1980	1151

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard Errors are presented in brackets.

Table 5: Descriptive Statistics: All Startups

	Star-Spawns						Other Spawns			Non-Spawns		
	More Similar to Parents			Less Similar to Parents								
	mean	st. dev	N	mean	st. dev	N	mean	st. dev	N	mean	st. dev	N
Secured External Financing	0.62	0.49	73	0.51	0.50	360	0.39	0.49	1332	0.23	0.42	4757
Initial Round Funded by												
Prominent Investors	0.71	0.46	45	0.63	0.48	183	0.48	0.50	514	0.35	0.48	1071
No. of Founders	1.89	0.91	73	1.78	0.70	360	1.61	0.66	1332	1.46	0.59	4757
Prior Founder Experience	0.44	0.50	73	0.36	0.48	360	0.46	0.50	1332	0.00	0.02	4757
Based in U.S.A.	0.86	0.35	73	0.81	0.39	360	0.69	0.46	1332	0.59	0.49	4757
Based in California	0.49	0.50	72	0.49	0.50	350	0.33	0.47	1315	0.21	0.40	4711
Based in Massachusettes	0.07	0.26	72	0.04	0.20	350	0.03	0.18	1315	0.03	0.17	4711
Based in New York	0.13	0.33	72	0.09	0.29	350	0.12	0.33	1315	0.09	0.29	4711
Founded in 2007	0.15	0.36	73	0.09	0.28	360	0.13	0.33	1332	0.12	0.33	4757
Founded in 2008	0.17	0.39	73	0.18	0.38	360	0.14	0.35	1332	0.18	0.39	4757
Founded in 2009	0.14	0.35	73	0.18	0.39	360	0.14	0.34	1332	0.14	0.35	4757
Founded in 2010	0.25	0.43	73	0.24	0.43	360	0.25	0.43	1332	0.21	0.41	4757
Founded in 2011	0.16	0.37	73	0.20	0.40	360	0.19	0.39	1332	0.19	0.39	4757
Founded in 2012	0.12	0.33	73	0.11	0.32	360	0.16	0.36	1332	0.15	0.36	4757

Table 6: Regression Coefficients: All Startups

	Secured External Financing	Initial Round Funded by Prominent Investors
	(1) Cox Proportional-Hazards	(2) Logistic
(a) Star-Spawn More Similar to Parents	1.266***	1.514***
	(0.228)	(0.476)
(b) Star-Spawn Less Similar to Parents	0.672***	0.633***
	(0.129)	(0.216)
(c) Non-Star Spawns with Tech Company Experience	0.358***	0.316
	(0.086)	(0.204)
No. of Founders	0.521***	0.168
	(0.049)	(0.111)
Prior Founder Experience	0.384***	0.299
	(0.098)	(0.213)
Industry Fixed Effects	Yes	Yes
Geography Fixed Effects	Yes	Yes
Founded Year Fixed Effects	Yes	Yes
p-value for Test: (a) = (b)	0.016	0.079
p-value for Test: (a) = (c)	0.000	0.015
χ^2	979.59	211.50
Log Likelihood	-9155.86	-1296.48
N	6131	1607

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard Errors are presented in brackets.

Table 7: Top 5 Words for Each Industry from Latent Dirichlet Allocation

Industry No.	Top 5 Words	% Sample
1	fashion, com, store, shopping, retailer	13.01
2	video, music, content, digital, entertainment	3.44
3	design, agency, seo, search, website	5.40
4	sports, network, dating, com, community	2.21
5	students, learning, job, education, training	3.02
6	users, share, allows, people, friends	6.38
7	companies, entrepreneurs, startup, startups, new	3.44
8	manufacturer, range, supplier, manufacturing, quality	2.33
9	energy, systems, power, develops, technologies	3.97
10	added, united, crunchbase, group, leading	3.31
11	wireless, communications, network, networks, provider	3.37
12	financial, real, insurance, investment, estate	2.71
13	health, healthcare, care, medical, law	2.45
14	businesses, small, email, payment, solution	4.18
15	data, cloud, security, enterprise, applications	5.28
16	search, com, local, marketplace, deals	4.15
17	food, home, cleaning, repair, best	2.27
18	medical, develops, developing, treatment, diseases	3.47
19	games, apps, app, game, applications	3.28
20	news, com, information, site, website	2.89
21	travel, event, events, hotel, hotels	2.54
22	people, make, quot, way, help	4.07
23	founded, new, san, york, california	1.94
24	consulting, provider, leading, global, companies	5.78
25	advertising, digital, brands, content, data	5.10