Who Finances Disparate Startups?

S. Katie Moon Paula Suh*

Abstract

Recently, new firm formations have become more geographically dispersed with greater regional industry diversity. Using detailed early-stage firm information from Crunchbase, we show that such a diminishing industrial agglomeration trend for young firms is driven by angel financing. This trend is tied to angel investors' unique portfolio selection of startups that diverges from venture capital's approach. Specifically, angels who are exceedingly intolerant of geographic distance prefer to invest in more distinctive firms industry-wise, while venture capital investors make industry-concentrated investments with relatively greater geographic flexibility. We also show that angel investors' portfolio selection of disparate startups enhances funded firms' performance and plays an important economic role in forming the regional entrepreneurial ecosystem.

Keywords: Firm Formations, Geographic Regions, Business Similarity, Angel Investment, Venture Capital Investment

JEL Classification: M13, L26

*Moon is at University of Colorado at Boulder, Leeds School of Business, Campus Box 419, Boulder, CO 80309, USA, [katie.moon@colorado.edu]. Suh is at the University of Georgia, Terry College of Business, 620 S. Lumpkin St., Athens, GA, 30602, [paula.suh@uga.edu]. The authors are grateful to seminar participants at the University of Colorado, University of Georgia, FOM Research Group Virtual Corporate Finance, and conference participants at the FMA 2021 meeting, the Colorado Finance Summit 2021, CICF 2022, as well as Hadiye Aslan (discussant), Tetyana Balyuk, John Barrios, Andrea Buffa, Maria Chaderina, Susan Chaplinsky, Michael Ewens, Jack He, Shenje Hshieh (discussant), Yael Hochberg, Pekka Honkanen, Steven Malliaris, Yifei Mao, Gordon Phillips, Shrihari Santosh, Antoinette Schoar, Ilya Strebulaev, Sheridan Titman, Heather Tookes, Yuri Tserlukevich, Xinxin Wang (discussant), Malcolm Wardlaw, and Xingtan Zhang. An earlier version of this paper was circulated under the title "Startup (Dis)similarity and Types of Early-stage Financing."

1 Introduction

Recent agglomeration literature has greatly expanded into entrepreneurial agglomeration based on the Marshallian spillovers.[1] Earlier studies highlight that the entrepreneurial agglomeration is highly localized and benefits from common input sharing, quality of matching in local labor markets, and knowledge spillovers (Jaffe, Trajtenberg, and Henderson, 1993; Rosenthal and Strange, 2008; Kolympiris, Kalaitzandonakes, and Miller, 2011). Entrepreneurial finance is also thought to play an important role in the agglomeration of entrepreneurship and innovation as spatial proximity helps with screening ventures, monitoring and advising portfolio firms, and thus mitigating information asymmetry and moral hazard (Gompers, Gornall, Kaplan, and Strebulaev, 2020; Gornall and Strebulaev, 2021; Hochberg, Ljungqvist, and Lu, 2007, 2010). These agglomeration benefits have persisted for some time, such that Silicon Valley has been the nation's leading technology hub for a long time.

However, a recent trend shows increasingly geographically dispersed formations of firms, especially in high-tech industries.[2] Panel A of Table 1 shows the list of the rising startup hubs by the firm growth from 2007 to 2018 based on our own data from Crunchbase. San Francisco and San Jose, the Metropolitan Statistical Areas (MSAs), which are considered Silicon Valley, are still ranked top in the list. What is surprising is that there are many new rising startup hubs, such as Austin, New York, and Boulder, which are growing as fast as San Jose, and other smaller cities such as Seattle, Durham, and San Diego, which are also often mentioned by the media as growing tech hubs. We also note that some areas that are not as often brought to the media attention (*e.g.,* Boise City-Nampa and Reno) are also on the list. This table confirms that startup formations have become increasingly more dispersed geographically.

[1]Marshall (1920) emphasizes that agglomeration ultimately reflects gains that reduce costs of moving goods, people, and ideas. See for example, Glaeser, Kerr, and Ponzetto (2010); Carlino and Kerr (2015); Chatterji, Glaeser, and Kerr (2014) among others.

[2]For example, see the Economist (September 1, 2018), "Silicon Valley is changing, and its lead over other tech hubs narrowing"; Kenan Insights report (June 1, 2022), "Capital Ecosystems: Seeding Smaller and Regional Funds to Increase Opportunity"; Bloomberg (August 2, 2018), "The Winners and Losers of America's Startup Economy."

[Insert Table 1 Here]

This recent geographic dispersion of startups is in part driven by high-tech firms' decreased reliance on the proximity to physical resources. For example, Silicon Valley's own products and services, such as cloud computing, video-conferencing, and online collaboration, have made it possible to exploit the traditional benefits of geographic spillovers of knowledge and labor in ventures anywhere, where living costs and competition are lower (The Economist 2018). More importantly, the difficulty of accessing traditional financing and, to a large extent venture capital (VC) financing,[3] raises the relative benefits of launching businesses in areas where other early-stage financing might be accessible. The idea of startup firms' seeking other types of early-stage financing is consistent with an interesting fact that we show in Panel B of Table 1 by comparing the average growth rates of firm formation and those of angel investors who are geographically spread over the nation. We find that the growing trend of firm formation in regions keeps its pace with increases in the number of local angel investors.

In this paper, we examine whether recent firm formations diverging from regional concentrations of similar industries are associated with angel investors. We offer an alternative perspective on a trade-off between the benefits of agglomeration and better access to financing. In doing so, we particularly highlight the rising role of angel investors, who are geographically spread over the nation and can be substitutes for venture capitalists (Hellmann, Schure, and Vo, 2021), in paving the way for the geographic dispersion and industry declustering of new startups.

The literature on industrial and geographical agglomeration is vast but with no definite resolution on the debate on spatial agglomeration as sources of the regional economic performance (see the survey by Beaudry and Schiffauerova (2009)). Marshall (1890), Arrow (1962), and Romer (1986) (the Marshall-Arrow-Romer model) argue that the concentration of an

[3]Puri and Zarutskie (2012) find that only 0.10% of firms per year have ever received VC financing on average over the period from 1981 to 2005.

industry in a region with a local monopoly promotes knowledge spillovers and facilitates the economic growth of the region, whereas Jacobs (1969) asserts that industry diversity in a region promotes innovative activity and economic growth as opposed to industry specialization. However, what is missing in the literature is the systematic mechanism for industrial and geographical agglomeration. Our paper fills the gap by showing that angel investors serve as a conduit for small-scale regional agglomeration with *industry diversity* by bridging disparate startups under their portfolios, which is consistent with Jacobs (1969).

We use the startup firm data from Crunchbase, the leading platform for superb information on early-stage firms for the sample period from 2007 to 2018. Particularly, Crunchbase tags firms with multiple industry classifications comprised of 47 broad industry groups and 742 detailed industries. Using these detailed industry classifications, we create an industry vector of 742 elements for each firm and compare those industry vectors across all firms within an MSA. To examine industry diversity in an MSA, we compute a cosine-similarity measure of these industry vectors. It is important to note that the similarity measure is time-varying as the geographic peers change over time due to the new entries or closures of firms. This is an important component of externality in this measure that later serves as one of our identification strategies for casual interpretations.

We first document stylized facts on the investment strategies of angel investors. By examining investor-level portfolios from the Crunchbase data, we find that angel investors' portfolio characteristics are significantly different from VC, which is another prominent financier of entrepreneurial financing. Specifically, angel investors' portfolios are more likely to be diverse across industries within a smaller geographic distance. This implies that angel investors tend to rely on physical proximity to alleviate information frictions between them and funded firms and by doing so they are more willing to fund diverse startup firms within close proximity.

Based on this finding, we examine whether strong presence of angel investors in an MSA is associated with an increase in business diversity in the region. We exploit the two

Securities and Exchange Commission (SEC) regulatory shocks to angel investments. We use Dodd-Frank amendment of the accredited investor definition in 2010 and the general solicitation amendment under the Jumpstart Our Business Startups (JOBS) Act in 2013 to show that increased angel investor participation in a region leads to higher local business diversity of startup firms. We further explore the mechanisms that can explain the positive association between active angel participation and business diversity by investigating startup firms' location choices. We expect startup firms that are open to angel financing to consider angel investors' investment strategies as an important factor in choosing their location to launch businesses. We find strong evidence supporting our prediction that a firm is more likely to enter into a market when it is dissimilar to the potential peers of the market given the relatively strong presence of local angel investors.

To give grounds for startup firms' incentives to differ from incumbent businesses in an area with strong presence of angel investors, we next examine how a firm's business similarity to geographic peers is associated with the type of early-stage financing that the firm receives. Our results indeed show that startup firms are more likely to receive angel financing when they are dissimilar to their geographic peers. A one standard deviation decrease in a firm's similarity to its geographic peers is related to doubling the unconditional likelihood of receiving angel financing. We find strikingly opposite results on VC funding that a decrease in business similarity *decreases* firms' access to VC financing. These contrasting results for angel vs. VC funding are consistent with the different investment styles of angel investors from VC investors and also effectively mitigate the concern about an unobserved regional economic shock that attracts more dissimilar firms and increases local funding opportunities at the same time.

Our main findings on the business dissimilarity and angel funding are robust to three alternative similarity measures. We find consistent results using the vintage similarity computed with only the initial set of 397 industry categories chosen by firms founded in 2007 to mitigate the concern about enlarging industry categories over time. Our results also hold for the soft-cosine similarity, which considers correlations among sub-classifications under a broad industry group, and for the dynamic similarity, which considers industry classifications dynamically updated by firms over time. The result is also robust to additionally supplementing angel investment data from the SEC Form-D filings. We take into account that angel financing and VC financing can be sequential for some firms and thus consider only the first funding rounds. We find our results are robust to using this first-funding sample. Also, we exclude firm-entry years to focus on the changes in a firm's business similarity after its entry and find consistent results. This result supports the causal interpretation of the effects because the changes in our measure of similarity to geographic peers in this case are beyond the firm's initial entry decision and exogenously given by the industry characteristics of newly entering or closing firms each year.

Next, we turn our focus to angel investors and examine whether specific characteristics of angel investors strengthen the relation between the business dissimilarity and angel funding. We are particularly interested in knowing how the local angel diversity affects their investment decisions to fund more diverse firms. We consider two dimensions of individual angel investors' diversity: Demography and education of local angel investors. From this analysis, we find that diverse startup firms are more likely to be funded by angel investors when those investors' demographic or educational backgrounds are more diverse in a region. This result has important policy implications in that individual diversity in a region can be a driving force to nurture unique business ideas among startup firms.

Our last analysis is on entrepreneurial outcomes by examining the successes of both angel investors and funded firms. Our results indicate that angel investors who fund firms that are more industry-diverse tend to have higher success rates as measured by exits through IPOs or acquisitions. Consistent with the idea that regional industry diversity allows angel investors to form a funding portfolio that improves the success of startups funded by those angels, we also find that firms targeting angel financing after entering into a region where they can stand out as a disparate business are more successful manifested in more subsequent fundings and the likelihood of an IPO or acquisition.

Overall, our results collectively support the idea that disparate startups are considered more favorably by angel investors. This mechanism may incentivize new entrepreneurs to start a unique business in a market with less industry clustering especially when angel investors' presence is greater in the market. Considering a significant increase in the number of angel investors (Figure 1), who are widespread throughout the country, we ascribe our evidence of geographic dispersion and industry declustering of new startups to spreading angel financing.

[**Insert Figure 1 Here**]

Our paper relates to three areas in the finance and economics literature. First, our paper adds to the entrepreneurial agglomeration literature. The recent agglomeration literature expands on the intellectual spillover with industry specialization (Marshall, 1890) and shows that entrepreneurial activities are much more localized than other economic forces linked to agglomeration (Rosenthal and Strange, 2008; Glaeser et al., 2010; Kolympiris et al., 2011; Carlino and Kerr, 2015; Chatterji et al., 2014). Particularly, VC has been extensively studied as an important factor of entrepreneurial agglomeration as VC tends to locate close to tech hubs and attract new startups in proximity (Chen, Gompers, Kovner, and Lerner, 2010; Hochberg et al., 2007, 2010). Chen and Ewens (2021) also examine regional startup agglomeration by focusing on local VC financing constraints as a mechanism and find that regionally disproportionate shocks to the supply of VC funds have spillovers effects on local startups' financial constraints and location choice. We expand the literature by linking the increasing trend of geographically dispersed new firm formations with the emergence of angel investors as the rising source of early-stage financing, which attracts startups away from more expensive and competitive tech hubs and trades off the traditional agglomeration benefits with angel funding opportunities. Especially, angel investors attract disparate startups into smaller regions and facilitate the different types of agglomeration benefits based on industry diversity, which is supported by Jacobs (1969). Our paper is the first to document the evidence that angel investments are one of the important mechanisms that drive industry diversity in geographically dispersed regions.

Second, our paper is closely related to the literature on angel investors. Studies on angel investors are relatively underdeveloped due to the data limitations on angel financing, compared to the extensive work on VC financing. Previous studies on angel investors focus on the real effects of angel financing and the success of angel-financed startups. Kerr, Lerner, and Schoar (2014) show that angel investments boost startups' survival and performance but not their likelihood of future fundraising. Lindsey and Stein (2019) show that a reduction in a market's pool of angel investors is negatively associated with firm entry and local employment. In contrast, Denes, Howell, Mezzanotti, Wang, and Xu (2020) find that angel investor tax credits have no significant effect on entrepreneurial activity. Hellmann et al. (2021) show that angels financing and VC financing are dynamic substitutes. We highlight that our study particularly focuses on angel funding for the startups *dissimilar* to incumbent firms in a spatial unit and is thus distinguished from the previous papers that examine the effects of angel financing on entrepreneurial outcomes of general startups. Our paper complements earlier studies of angel investors by showing that the local startup business diversity is an important factor in angels' funding decisions (Bernstein, Korteweg, and Laws, 2017) and successful exits (Kerr et al., 2014; Lerner, Schoar, Sokolinski, and Wilson, 2018).

Lastly, our paper broadly connects to the literature on early-stage financing and financial contracting. We consider the preferences of angel investors in creating and diversifying their investment portfolios and show that the diversification motive of angel investors with respect to the industry is closely related to regional industry diversity. These investment preferences of angel investors, summarized as geographic concentration and industry diversification, are contrary to those of VC investors who prefer industry specialization (Sorenson and Stuart (2001); Hochberg et al. (2007); Chen, Gompers, Kovner, and Lerner (2009)). The difference in investment preferences highlights how those different types of investors juggle the two-dimensional information frictions through physical distance and knowledge gap.

2 Data and Sample

2.1 Crunchbase Data

We collect our main data on startup firms and their funding information from Crunchbase. Crunchbase is a crowdsource platform started in 2007 and provides information about private and public companies and investors. Its parent company is TechCrunch which is an online newspaper focusing on high-tech and startup companies. TechCrunch originally operated Crunchbase as its online encyclopedia for startup company information. Although the initial data relied on web searching and scraping, now Crunchbase collects data directly from investment firms that submit monthly portfolio updates to Crunchbase and its wide network of partners covering companies, executives, entrepreneurs, and investors. Furthermore, Crunchbase maintains high quality data through machine learning algorithms to validate data for accuracy and anomalies daily and ensures capturing notable funding rounds, acquisitions, and exits by following over 2,000 of the top news publications. Lastly, Crunchbase covers the most comprehensive set of startup firms, especially high-tech startups. The geographic coverage is not just restricted to the well-known innovation hub in California as shown in Internet Appendix Table IA.1.

Crunchbase data has been used by academics in finance for research in early-stage financing. Kaplan and Lerner (2016) introduce Crunchbase as the best known source of venture capital financing. Further, the use of the data has expanded since covering details of all types of early-stage financing beyond venture capital financing. Davis, Morse, and Wang (2020) use detailed startup financing types, venture debt in particular, from Crunchbase to study startup firms' capital structure decisions and successes. Ewens, Nanda, and Rhodes-Kropf (2018) use the startup team member characteristics for examining VC portfolio strategies. Wang (2018) uses the Crunchbase data for tracking entrepreneurship decisions. In this paper, we make extensive use of Crunchbase for the detailed industry classifications to create our unique business similarity measure. We also benefit from the detailed location information of firms and investors and funding information by different investor types from Crunchbase.

We access the snapshot of information on 753,938 organizations on Crunchbase as of April 2019 with unique Crunchbase identifiers (uuid). The organizations are further categorized into company, investor, and school. We keep organizations whose primary role is "company." Each company receives a unique Crunchbase identifier (uuid), and all information such as investors and funding rounds is categorized into a node that also receives a unique identifier. We utilize these identifiers to connect firms to their funding investors and rounds information. Besides, Crunchbase also provides detailed information on exits including IPOs and acquisitions and founder and team member information.

Despite the fact that Crunchbase provides the most comprehensive data on early-stage startup companies, one may be concerned with the possibility that firms on Crunchbase are systematically different (i.e., self-select to be listed on Crunchbase). However, for this selection issue to explain our main results, the bias has to be correlated both with the business similarity and financing choice. This seems unlikely. First, the similarity measure is a dynamic measure affected not only by the firm's choice but also by the neighboring firms' choices. Hence, a firm's choice to be on Crunchbase does not seem to correlate with our similarity measure in any clear way. Second, being on Crunchbase does not guarantee access to financing or any specific type of financing. We find that 18% of firms in our sample have at least one funding record, and the funding type covers a wide spectrum from seed money to post-IPO debt financing.

2.2 Sample Selection

We restrict our sample to firms that have non-missing headquarter location, valid zip code, and founding year. For the sample of startups, we limit their age up to ten years since their founding year.[4] It is important to note that Crunchbase includes not only startups but also

[4]Some private firms in our sample are very old. For example, Scovill Fasteners was founded in 1802, and thus skews the age distribution to the right. Hence, we use the median age at the IPO of 2,067 exiting firms in our sample, which is ten, as the age cutoff.

public firms. Geographic peers to startup firms in our analyses include those public firms, and we do not screen them based on the age cutoff when calculating the business similarity. We exclude observations from the sample after firms close their business. After all these screens, our final sample comes down to 119,605 unique firms founded between 2007 and 2018.[5] We extend the data with investor and funding information, which results in about 1.5 million firm-year observations. About 18% of the sample firms have at least one funding record (denoted funding sample) and corresponding funding investor information.

Table 2 describes firm age and funding information for the funding sample. The average (median) age of firms in our funding sample is 5.5 (5). Firms in the funding sample have on average about two funding records. The average age at which firms receive funding is 3.07, whereas the age at receiving the first funding record is slightly younger at 2.48.

[**Insert Table 2 Here**]

We then summarize funding data from angel and VC investors separately in detail. A smaller fraction of firms in the funding sample receive funding from angel investors than from VCs (19% vs. 59%). Firms are younger (2.35 vs. 3.37 years old) at the time they receive angel financing than VC financing. The statistics on funding amounts and the geographical distance between funded firms and corresponding investors are consistent with earlier studies (MIT Entrepreneurship Center, 2000). The median size of angel financing is smaller than that of VC financing ($875,000 vs. $4.3 million). Angel investors tend to fund firms that are in significantly closer geographic proximity than do VC investors (88 miles vs. 427 miles).

[5] Although Crunchbase provides information on founding activities even before 1990, we intentionally restrict our sample to the period after 2007 when Crunchbase launched in July of 2007. Our results are robust to including the data before 2007.

3 Business Similarity

3.1 Business Similarity Measure

One unique feature of Crunchbase is its industry classification of each firm. Since a large proportion of Crunchbase firms are startups, they do not use conventional industry classifications such as SIC or NAICS code. Instead, Crunchbase has created more technology-oriented industry classifications of its own. There are 46 broad classification, which covers anything from Consumer Goods to Information Technology, and total 742 finer sub-classifications. See Internet Appendix B for the 742 industries and their groups. For example, Information Technology further breaks down into Cloud Data Services, Cyber Security, Data Integration, Sales Automation, Video Conferencing, and others. Thus, Crunchbase industry classifications provide very detailed industry coverage from conventional industries, such as consumer goods, to high-technology industries that better capture a wide variety of business within startup companies. Firms can have multiple categories that describe their business at the time of creating a company profile and are allowed to update them over time.[6] In our sample, firms on average report three categories (median of 2.73). Only about 10% of firms report more than 5 categories.

Using this big advantage of 742 detailed industry classifications provided by Crunchbase, we create a business similarity measure for each firm-year. We first create a vector of 742 elements for each firm where each element corresponds to one of the 742 industries. Each element receives one for the reported category and zero otherwise. Then, we use these vectors to compute the cosine similarity score of a pair of a given firm and its geographic peer. Suppose that there are N firms in an MSA in a year. A given firm i's industry vector can be represented by a vector of v_i with element h ($1 \leq h \leq 742$) being one if firm i's industry classifications by Crunchbase include the category h and zero otherwise. Then, we

[6] In Internet Appendix Table IA.2, we provide a list of the top 20 popular categories chosen by firms over time. The list seems to ensure that the Crunchbase categories depict industry trends in a timely manner. For example, whereas Software has always been the most popular category, Artificial Intelligence, Blockchain, FinTech, and Machine Learning have become rising categories in most recent years.

compute the firm-to-firm business similarity using the vectors of v_i and v_j for a pair of firm i and firm j based on the product cosine similarity:

$$Pairwise\ Business\ Similarity_{i,j} = (v_i \cdot v_j). \tag{1}$$

The pairwise business similarity ranges from zero to one. We end up with $N-1$ pairwise business similarity scores for each firm and then take the average of them to denote a firm's similarity to its geographic peers within MSA for a given year. We compute the similarity scores for each startup company from 2007, the year in which Crunchbase was founded, but consider all firms founded before 2007 as existing geographic peers. As new firms are added to an MSA, the similarity score evolves over time for both the existing and new firms in the data. The economic interpretation of the similarity score is that it serves as a proxy for how close a firm is to its geographic peers within MSA in terms of Crunchbase industry classifications; higher values of similarity suggest that the firm clusters closely with its industry peers within MSA.

The average of similarity scores during our sample period is 0.0384 with a standard deviation of 0.0330 from approximately 1.5 million firm-year observations. We illustrate magnitudes of similarity scores and their variation with a few notable examples to help with the interpretation of our results. MSAs that have similarity scores close to the sample median are Columbus, Ohio (0.0330), Fort Wayne, Indiana (0.0331), and New York-Northern New Jersey-Long Island, NY-NJ-PA (0.0331). The most intuitive comparison would be Silicon Valley which belongs to the San Jose-Sunnyvale-Santa Clara MSA and whose similarity is 0.05. As expected, the Silicon Valley area's similarity score is above the sample average as it is comprised of similar technology-based firms. A one standard deviation increase of 0.0330 in similarity scores would then corresponds to entering Silicon Valley compared to Modesto, California (0.0175). Modesto is about 80 miles northeast of Silicon Valley with fertile farmland and the largest winery in the world. The industries in Modesto are more diverse than Silicon Valley ranging from winery and food producers to bottle manufacturing and steelworking companies.[7]

4 Angel Investors and Identification Strategies

4.1 Characteristics of Angel Investors

Prior studies document stylized facts on angel investors in contrast to VCs. Wharton Entrepreneurship and Angel Capital Association (2017) reports that angel investors are geographically dispersed relative to VCs and cluster in a few cities: 63% of angel investors are located outside California, New York, and Boston, with a sizable presence across Great Lakes, Southeast, and Mid-Atlantic. MIT Entrepreneurship Center (2000) and Kauffman Foundation (2002) show that angel investors tend to invest in startups in close geographic proximity to their homes: most active angel investors investigate opportunities and memberships in their local areas and do not invest in opportunities outside a 1-2 hours driving distance from home. In contrast, the bigger investment size and the network of syndicates allow VCs to specialize in an industry (Hochberg et al., 2010; Hochberg, Mazzeo, and McDevitt, 2015; Lerner, 2020) while investing more frequently in spatially distant companies as the social network of VCs diffuses information across boundaries and expand spatial radius (Sorenson and Stuart, 2001; Hochberg et al., 2007). Furthermore, VC investors are less risk-averse than angel investors because the scale of VC investments makes other control mechanisms accessible to VCs. VCs can actively mitigate the information asymmetry using convertible securities, syndication, and staging (Gompers, 1995), contractual provisions (Kaplan and Stromberg, 2001), and board seats (Lerner, 1995). Wong (2002) finds that angel

[7] Other areas that have a similarity score comparable to Silicon Valley include Ann Arbor, Michigan (0.0506), Gainesville, Florida (0.0505), and Spartanburg, South Carolina (0.0493). Texarkana, Texas has the highest similarity score of 0.2954 and is known for the concentration of Tires, Wood products, Food and paper industries with facilities of large firms (*e.g.,* Alcoa, Cooper Tire, Rubber, and Red River Army Depot) in the area. Clarksville, Tennessee-Kentucky (0.0076), Lake Charles, Louisiana (0.0069), and Sheboygan, Wisconsin (0.0067) are the areas with the lowest similarity scores.

investors are often not involved in the subsequent rounds in staged financing, only 0.59 board seats are added in angel rounds compared to 1.12 seats in VC rounds, and some contractual provisions, such as first refusal provisions, contingent equity stakes, and full ratcheting protections, are less common in angel funding.

Therefore, we expect angel investors to have different mechanisms to alleviate the underlying information frictions than VC investors, despite being exposed to the same underlying information frictions. We simplify potential information frictions into two dimensions, physical distance vs. knowledge distance. The geographic distance makes monitoring and advising more difficult, and the knowledge gap makes the investment less informed, both of which exacerbate information asymmetry. We provide a theoretical framework in Internet Appendix A that highlights that different types of early-stage financiers have different mechanisms to alleviate the underlying information frictions. We then examine in Table 3 over which dimension angel investors are more likely to minimize the information asymmetry compared to other types of investors.

[Insert Table 3 Here]

In columns (1) and (2) of Panel A, we consider the business similarity of funded firms in each investor's funding portfolio. The funded firm business similarity in a portfolio is the average firm business similarity to the rest of the firms in the portfolio. This test is cross-sectional across investors, and thus the sample consists of one observation per investor. If an investor has only one funded firm in the portfolio, we drop the investor from the analysis because the business similarity with other funded firms can not be computed in this case. In column (1), we consider all three types of investors, angel, VC, and other investors with one of the remaining types as a control group. In column (2), we only compare angel investors with VCs as a control group.

Column (1) shows that VCs have a strong preference to fund companies with similar industry classifications relative to both the angel and other types of investors. In column (2) when we compare angel investors and VCs, we find that funded firms in angel investors' portfolios are relatively more dissimilar than those in VC portfolios. The coefficient estimate for *Angel* in column (2) indicates that the average business similarity score for angel investors' portfolios is lower than VC portfolios by 1.8 percentage points or 19.4% from the unconditional mean of portfolio-firm similarity at 0.093. We repeat analogous tests in columns (3) and (4) by replacing the dependent variable with the average geographical distance between a given investor and the funded firms in its funding portfolio. Both columns (3) and (4) show that angel investors have a significantly stronger preference to maintain geographical proximity to their funded firms than that of VCs or other types of investors. For example, the coefficient estimate for *Angel* in column (4) implies that the average distance between angel investors and their portfolio firms is about 107 miles closer, or 14.7% shorter compared to the mean distance between VCs and their portfolio firms of 725 miles.

In Panel B of Table 3, we consider funding-level observations instead to examine business similarity and geographic distance between existing portfolio firms and newly funded firms. Consistent with the results in Panel A, we find in columns (1) and (2) that angel investors strongly prefer to add dissimilar firms in their funding portfolios relative to VCs and other types of investors. Similarly, we continue to find in columns (3) and (4) that angel investors fund new firms when those firms are in close proximity to the existing funded firms in their portfolios. For example, in column (4) angel investors choose to invest in a firm that is approximately 52 miles closer to the funded firms in their existing portfolios relative to VC investors. The results in Table 3 collectively show that angel investors are more likely to rely on physical proximity to alleviate information frictions between them and funded firms while VCs tend to resort to their industry expertise.[8]

Lastly, we show that the choice of dimension in overcoming information frictions is not

[8]We also find that angel investors and VCs are significantly different in investment frequencies. Internet Appendix Figure IA.1 (left panel) shows that during our sample period, the median number of investments made by angel investors is about two investments every ten years, while VC investors have made more frequent investments at six every ten years. This indicates that when angel investors make their funding decisions, it is likely that they comprehensively consider the project characteristics including geographic distance and business similarity over the years rather than promptly evaluating the flow of new deals as they come in as VC investors are known to do so.

a random characterization of angel investors and VCs but rather a function of investment size. Unlike individual investors who are small in scale and lack subject matter expertise, as an investment size grows, financiers' understanding of technology, market, and people can improve and their subject matter expertise can accumulate. Hence, we expect angel investors' strong preference for reducing the physical distance over the knowledge gap decreases with the investment size. We test this hypothesis using the wide spectrum of our angel investor sample from a single person to a large group of angels, who resemble micro-VCs. Since VCs are known to specialize in an industry, we only use the angel investor sample to test for the hypothesis. In Figure 2, we plot the trade-off between reducing the geographic distance and increasing industry specialization by angel investment sizes (represented by color). We find that the angel investors' preference for reducing the geographic distance over industry diversity decreases monotonically with the angel investment size and increasingly resembles the VCs' preference for industry specialization when the angel investment size grows sufficiently large.[9] This shift is consistent with the greater rent extraction and resource reallocation efficiency benefits of large investments for more specialized portfolios shown by Fulghieri and Sevilir (2009).

[Insert Figure 2 Here]

Overall, results in Table 3 and Figure 2 show strong evidence that angel investors maintain geographic proximity and hold more dissimilar startups in their funding portfolio. The evidence in this section supports the conclusion that business similarity plays an important role in determining the types of early-stage financing and that angel financing is likely associated with the increased business diversity in entrepreneurship in a region.[10]

[9]When the angel investment size is too small (brightest yellow plot), i.e. only two funded firms in a portfolio, the diversification benefit may be limited.

[10]We further describe angel investors from our sample more in detail in Internet Appendix Table IA.3. There are 7,125 unique angel investors that have funded firms in our sample. The angel investors make on average 3.69 investments during our sample period. Angel investors are dominated by male investors and hold on average 1.51 academic degrees with 24% and 5% of them holding MBA and PhD degrees, respectively. About half of the angel investors (i.e., 3,279 out of 7,125) have founded at least one company, and conditional on having entrepreneurial experience, they have founded on average 1.67 companies. Also, angel investors hold on average about two advising roles for entities.

4.2 Identification Strategies

In this paper, we aim to show causal effects of angel investors on local startup dynamics. Specifically, we consider the two SEC regulatory shocks to angel investment based on the Dodd-Frank amendment of the accredited investor definition in 2010 and the general solicitation amendment under the Jumpstart Our Business Startups (JOBS) Act in 2013. Both rules changed the supply of angel funding and subsequently affected local startups' angel financing.

Startups rely on Rule 506 of Regulation D, which is a safe harbor based on the exemption provided in Section 4(2) of the Securities Act for sales of private (unregistered) securities sold to accredited investors. The sale of private securities sold to accredited investors requires no disclosure from the issuer and puts no limit on the dollar amount of the offering and the number of sales to accredited investors. Angel investors, who fund startups through private securities, are often considered as accredited investors. For these private securities to be exempt from similar regulatory requirements as public securities have, the issuers of private securities must not use general solicitation to market the securities (*i.e.,* the prohibition against general solicitation and advertising in Rule 506 and Rule 144A Offerings).

As the first regulatory shock, we exploit the changes in the definition of accredited investors introduced by the Dodd-Frank. To qualify as an accredited investor, the individual must have income in excess of $200,000 (or $300,000 joint-income for married couple) or a net worth over $1 million. The Dodd-Frank amended the rule that the value of a person's primary residence should be excluded from the calculation of net worth, resulting in the decrease in the number of net-worth qualified accredited investors.[11] We follow Lindsey and Stein (2019) in defining our marginal treatment group as the fraction of investors who may have lost their accreditation after the Dodd-Frank at the state-level. Using the income and net worth data from the Survey of Income and Program Participation (SIPP),

[11] The amendments were further updated to reflect that positive home equity should not be included in the calculation of net worth and imposed restrictions on the use of cash-out mortgage refinancing to meet the threshold.

a household-level longitudinal survey covering individuals and families, we create the percentage of families that likely have lost their accreditation after the Dodd-Frank.[12] Similar to Lindsey and Stein (2019), 2.6% families lost their accreditation due to the change made in the asset standard.

As the second regulatory shock, we consider the JOBS Act eliminating the prohibition against general solicitation in private offerings. The amendment permits an issuer to engage in general solicitation or general advertising, provided that all purchasers are accredited investors. Permitting general solicitation in private offerings means advertising an active capital raise to a broad audience, and thus effectively encourages general individual accredited investors' participation in private capital raising that would have been limited to particular groups of investors. According to the Director of the Division of Corporation Finances at the SEC, over 900 new offerings were conducted and more than $10 billion in new capital were raised relying on this amendment within only the six-months after the exemption became available (Higgins, 2014).

The two regulatory shocks we use are complementary as both of them change the angel investor participation in funding startups but in the opposite direction. The Dodd-Frank shock represents a *reduction* in the accredited investors, as measured by the percentage of families who lost accreditation in a region, whereas the JOBS Act shock represents an *increase* in the broad participation of accredited investors, as measured by the size of accredited investor pool in a region that could have been affected by the change. By exploiting these two regulatory shocks to angel investor participation, we examine the possible causal relation between angel investors and the industry diversity of startup firms.

[12]We use the 2008 SIPP Wave 10 panel conducted between September and December of 2011 since this is the only panel that contains special topical module with detailed questions about family income, asset, and liabilities for the assessment of accreditation status change under Dodd-Frank rules.

5 Business Similarity and Financing

5.1 Business Similarity and Angel Investors

Based on the conceptual framework and the empirical evidence from our data in Section 4.1, we predict that strong presence of angel investors will be positively related to the business diversity in the region as they fund more industry-diverse startups. In this section, we test for this direct relation first. We recognize possible endogeneity concerns for this relation. For example, regional economic prosperity can simultaneously affect the number of angel investors and the diversity of entrepreneurship by attracting both individuals and firms to the region. We attempt to mitigate this endogeneity concern using the two regulatory shocks only to the number or activities of regional angel investors.

In Table 4, we report results from the two difference-in-differences tests where we regress business similarity at the MSA-year level on the indicators for treated regions and periods after each regulatory shock. The regulatory shocks to angel investors are the Dodd-Frank amendment of the accredited investor definition in column (1) and the general solicitation amendment under the JOBS Act in column (2). Treated for the Dodd-Frank amendment in column (1) is one if the fraction of families that may have lost accreditation due to the regulation change at the state level is below median and zero otherwise. Post in column (1) is one after the enactment of Dodd-Frank in 2010 and zero otherwise. The sample in column (2) only includes the post-Dodd-Frank period after 2010 to avoid comingled effects from both regulations. Treated for the JOBS Act's general solicitation amendment in column (2) is one if the number of accredited investors at the family level during the pre-elimination period is greater than the median and zero otherwise. Post in column (2) one after the effective date of the general solicitation amendment in 2013 and zero otherwise. We specifically estimate the following regression:

$$Similarity_{m,t} = \alpha + \beta_1 Treated_m \times Post_t + \beta_2 Treated_m + \beta_3 Post_t + \gamma\Gamma_{m,t-1} + \epsilon_{m,t}, \quad (2)$$

where $\Gamma_{m,t}$ is a set of MSA-level control variables in the year prior to t, including the number of incumbent firms, number of closed firms, number of investors, population, and GDP. We include MSA and year fixed effects that subsume the estimations of standalone $Treated$ and $Post$. Standard errors are clustered at the MSA-year level.

[Insert Table 4 Here]

We find in column (1) that the treated states, which are less affected by the Dodd-Frank amendment and thus would experience a smaller decrease in the number of accredited investors than more affected states, have relatively more business diversity measured by a decrease in similarity scores after the rule change. The result with the JOBS Act's general solicitation amendment in column (2) is consistent. We find that the treated states, which are more affected by the permitted general solicitation and thus would experience increased investment activities by general individual investors, have relatively more business diversity after the rule change. The results in this table imply that local angel investors in both their number and active participation are strongly associated with the region's business diversity of startup firms. In the subsequent sections, we dig deeper this point and further examine startup firms' entry decisions and financing.

5.2 Business Similarity and Entry Decision

Thus far, our results show that more active angel investors in a region lead to business diversity of startups in the region. We now explore the mechanisms that can explain this association by focusing on startup firms' entry decisions. Our previous analysis in Table 3 shows that angel investors prefer more dissimilar startups in their funding portfolio while maintaining geographic distance. Thus, we hypothesize that startup firms aiming to receive angel funding will consider such preference of angel investors when founding their business and choose a location where their business appears to be more diverse compared to incumbent firms.

To explore firms' entry decisions, we consider cross-sectional data at the time of the entry year t and compute a firm's similarity to incumbents in each of 416 MSAs in the year prior to the entry. Hence, each firm has 416 cross-sectional observations. To mitigate potential concerns for endogeneity, we also employ DID test designs similar to Table 4 and additionally interact the DID estimates with an entering firm's potential business similarity scores to all incumbent firms. We specifically estimate the following regression specification:

$$\begin{aligned} Entry_{i,m} = \alpha &+ \beta_1 Dissimilarity_{i,m} \times Treated_m \times Post + \beta_2 Treated_m \times Post \\ &+ \beta_2 Dissimilarity_{i,m} \times Treated_m + \beta_3 Dissimilarity_{i,m} \times Post \\ &+ Dissimilarity_{i,m} + Treated_m + Post + \gamma\Gamma_m + \epsilon_{i,m}. \end{aligned} \tag{3}$$

The dependent variable, $Entry_{i,m}$, is one if MSA m is chosen by entering firm i at its entry year and zero otherwise. We then compute firm i's potential business similarity scores to all incumbent firms in MSA m at the entry year. For a pair of firm i and MSA m, we take the average of $Pairwise\ Business\ Similarity_{i,j}$ defined in Equation (1) where v_i is the industry vector of firm i and v_j is that of an incumbent firm j in MSA m. We denote the average business similarity for pairs of firm i and all incumbent firms in MSA m by $Similarity_{i,m}$. For ease of interpretation, we use $Dissimilarity$ henceforth by subtracting $Similarity$ from one. Incumbent firm data are lagged one year relative to firm i's entry year. Γ_m is a set of MSA-level control variables in the year prior to the entry year, including the number of incumbent firms, number of investors, and per capita personal income and GDP of an MSA from Bureau of Economic Analysis. We additionally control for local innovation ecosystem dynamics measures from Andrews, Fazio, Guzman, Liu, and Stern (2019) although the data end in 2013 and thus cannot be used for the general-solicitation analysis. We include MSA and entry-year fixed effects separately. Standard errors are clustered at the firm level. The results are presented in Table 5. To conserve space, we report only the estimated coefficients for $Dissimilarity \times Treated \times Post$ (the main variable of interest), $Treated \times Post$, and

Dissimilarity along with control variables.

[Insert Table 5 Here]

Column (1) shows that the triple interaction term of *Dissimilarity*, *Treated*, and *Post* is positive and statistically significant. This result indicates that startup firms are more likely to enter an MSA if their business appears to be more dissimilar to those of the incumbent firms, particularly when the MSA is in a treated state that is less affected by the Dodd-Frank amendment after the regulatory shock. In other words, startup firms prefer an MSA with more angel investors in choosing a location to launch its business. This effect is economically significant. A one standard deviation decrease in similarity (0.037) at the mean Dodd-Frank treatment state, where 37% of previously accredited investors lost accreditation, is associated with a 15% increase from the unconditional probability of 0.24 percentage points.[13] Besides the triple interaction term, we find that standalone *Dissimilarity*, number of investors, and per capita personal income and GDP of an MSA is also positively associated with firm entry. We also find that the total number of incumbent firms is negatively associated with firm entry, which implies that startup firms avoid overall local competition. The results with the JOBS Act's general solicitation amendment in column (2) are similar. A one standard deviation decrease in the similarity together with one standard deviation increase in accredited investors (76) after the amendment increases the probability of entry by 14.7% from the unconditional probability of entry in a random MSA.

Overall, the results in Table 5 support the conclusion that business dissimilarity plays a significant role in firm entry decisions when the presence of angel investors is relatively strong. In the next section, we further examine whether business dissimilarity indeed enhances the likelihood of receiving angel financing to give grounds for startup firms' incentives to differ from incumbent businesses in an area with strong presence of angel investors.

[13]The probability of a random choice of one MSA out of the total 416 MSAs is 0.0024. The standard deviation of *Similarity* in this analysis, 0.037, is different from 0.033 from our unconditional sample because we consider all potential similarity scores of a firm to all incumbent firms in both chosen and unchosen MSA locations for this analysis.

5.3 Business Similarity and Angel Funding

In this section, we examine how business similarity affects startup firms' access to angel financing. Each investor (whether an organization or a person) in Crunchbase reports one or more investor types. We group the 22 detailed investor types into three groups broadly: Angel, venture capital, and others. Angels include fund providers of "pre-seed", "seed", and "angel" investments, and VCs include fund providers of any serial rounds. Others include investment banks, hedge funds, pension funds, private equities, and accelerators. Based on these investor classifications, we regress angel funding on the business diversity. Specifically, we consider the following regression specification:

$$Angel\ Funding_{i,t} = \alpha + \beta Dissimilarity_{i,m,t} + \gamma\Gamma_{m,t} + \epsilon_{i,t}, \tag{4}$$

where $Angel\ Funding_{i,t}$ is either an indicator variable for funding received from angel investors in a year or the total amount of funding in million dollars received from angel investors in a year. The main variable of interest is $Dissimilarity_{i,m,t}$, which is one minus the average business similarity for pairs of firm i and all incumbent firms in MSA m. The specification includes the log number of firm age and log number of firms, G Index and EG Index for MSA m in year t as control variables ($\Gamma_{m,t}$).[14] Firm and year fixed effects are also included, and standard errors are clustered by MSAs. Table 6 presents the results.

[Insert Table 6 Here]

In columns (1) and (2), we find that an increase in business dissimilarity increases a firm's access to angel financing. The coefficient estimate for *Dissimilarity* in column (1), for example, imply that a one standard deviation (0.033) drop in similarity scores increases the

[14]G and EG Index are computed following Ellison and Glaeser (1997) using our data. The indices are measures of the geographic concentration of an industry, based on a location choice model considering the localized industry-specific spillovers, natural advantages, and pure random chance of plant location choices. The G Index is computed as the sum of squares of the difference between the observed concentration of state-industry employment beyond the model estimate, and the EG index further controls for the differences in the size of the distribution of plants and the size of the geographic areas.

probability of receiving funding from angel investors by 1.2 percentage points, equivalent to doubling the likelihood of receiving angel funding from the unconditional mean. The results are considerably larger for funding sample, where we only use firms that have received at least one funding during our sample period (Table IA.4). A one standard deviation drop in similarity scores increases the likelihood of receiving and the amount of angel funding by 23 and 13.6 percentage points, respectively.

Next, we consider three alternative measures of similarity in the remaining columns, including vintage similarity, soft-cosine similarity, and dynamic similarity. First, in columns (3) and (4), we consider the vintage similarity which is a cosine similarity measure using only the initial set of 397 industry categories from Crunchbase that are chosen by only firms founded in 2007, the start of our sample period. This is to rule out the possibility that potentially enlarging industry categories over time drive a decrease in similarity and our results. That is, industry categories that are newly created in the middle of our sample period such as "Blockchain" and "FinTech" are completely dropped for the analysis. Second, in columns (5) and (6), we replace our primary similarity measure with a more advanced version of similarity scores that additionally take into account correlations among finer sub-classifications within 46 broad industry groups, based on the soft-cosine similarity calculation technique introduced by Sidorov, Gelbukh, Gómez-Adorno, and Pinto (2014). Instead of treating all 742 elements in an industry vector as completely unrelated, we adjust the elements under the same broad industry group to be treated equally more similar.[15] Third, in columns (7) and (8), we consider potential changes of industry vectors within firm over time. It is possible that firms may voluntarily update their industry classifications as their businesses are expanding, contracting, or pivoting. To address this possibility, we use multiple Crunchbase data dumps that have been acquired in different years including 2016, 2018, 2019, and 2020 and create industry vectors that change over time.[16]

[15]The soft-cosine similarity technique is used in machine learning and natural language processing where text sentences are expressed on a non-orthogonal basis. For example, "Hi" and "Hello" are synonyms, and thus the angle between two basis vectors of "Hi" and "Hello" is set to be non-orthogonal.

[16]We thank Ilya Strebulaev for graciously sharing the Crunchbase data dumps in earlier years. When

The results in all six columns using alternative similarity measures show that our main findings on angel funding are consistent when we restrict industry classifications to the vintage year choices, consider correlations among sub-classifications under a broad industry group, or dynamically update reported industry classifications by firms over time. Specifically, the results using the vintage similarity in columns (3) and (4) are almost identical to those in the first two columns in terms of economic or statistical significance. The results using both soft-cosine similarity and dynamic similarity in columns (5) to (8) are similar to those in the first two columns, although the magnitude of the effect is weaker as about one second to third.

The sample that we use for this analysis consists of all firm-year observations in the Crunchbase data regardless of whether a firm receives any funding during the entire sample period. We note that only about 18% of the firms in our sample have at least one funding record and corresponding funding investor information. To examine early-stage financing in the full sample, we thus replace observations with no funding data from the Crunchbase with zeros. Accordingly, the effects we document thus far are more likely to capture extensive margins. In Internet Appendix Table IA.4, we consider the intensive margin effects by using firm-year observations with any reported funding (i.e., conditional on receiving at least one funding). We find that results are consistent overall while the intensive margin effects are weaker particularly for the dynamic similarity measure.

5.4 Robustness

There can be a concern specifically about an entry of a group of firms with certain firm characteristics attracting other dissimilar businesses into the same MSA and simultaneously increasing their own chance of obtaining angel funding relative to VC funding. Although such potential firm-level characteristics that cast a broad impact over the entire MSA are

there is no Crunchbase data dump for a specific year, we use the data from the closest prior year (e.g., using the data in the 2016 dump for the 2017 industry vectors). For the years that have no prior-year data, we use the earliest possible data (i.e., using the data in the 2016 dump for years from 2007 to 2015).

hardly known in this context (a potentially omitted firm variable that attracts only dissimilar peers and increases the likelihood of receiving angel financing for themselves), we specifically mitigate such concern in Panel A of Table 7 by excluding all entry year observations and allowing the results to be driven only by the subsequent external disturbances in the similarity scores.

[Insert Table 7 Here]

We find in Panel A that our results remain robust and consistent. For example, in column (1) of Panel A, we find that a one standard deviation increase in a firm's dissimilarity score, purely driven by other newly entering or closing firms in the same MSA, leads to an increase of the likelihood receiving angel funding by 1.3 percentage points which is almost identical to the results in column (1) of Table 6. In Figure 3, we additionally examine how a firm's similarity score evolves on average after its entry to examine the economic significance of the analysis in Panel A of Table 7. The box plot in Figure 3 shows estimated variations of similarity scores over firm age after they enter an MSA. We find in the figure that the variation of similarity scores is biggest in the year following the firm entry. The maximum variation is estimated at around 6% (0.0022) from the mean similarity score (0.0384) at the firm age of one. The variations appear to decrease over the firm age but remain at about 2% even at age of 10. Hence, the magnitudes of the post-entry variations in similarity scores from other newly entering or closing firms in the same MSA are economically significant and large enough to lead to meaningful external disturbances to the similarity scores.

[Insert Figure 3 Here]

There can be another concern that angel funding and VC funding are sequential for some firms and thus VC funding generally follows angel funding. To address this concern, we examine how business similarity affects the type of the first funding by restricting the sample to only the very first funding observations. We repeat the regression in column (1) of Table 6 with the first-funding sample and report the results in Panel B of Table 7.

The dependent variable in Panel B is an indicator variables for the first funding that is received from angel investors. Since each firm has only one first funding observation, we include various other fixed effects in place of firm and year fixed effects. Those alternative fixed effects are funding year and founding year fixed effects, funded firms' MSA and investors' MSA fixed effects. In all columns except for column (3) that uses soft cosine similarity, we find that the coefficient estimates for *Dissimilarity* are positive and significant in predicting angel funding as the initial funding of a firm. These results indicate that a lower business similarity score of a firm to its local peers increases the probability that the first funding is received from angel investors. The magnitudes of the effect are slightly larger with this initial funding analysis than those in Table 6.

One may be still concerned about some unobserved economic shock to an MSA that attracts new startups, resulting in a local composition of more dissimilar firms, and increases the local funding opportunities at the same time. However, the essence of our results is that any external shocks that disturb the business similarity of an MSA including changes in economic conditions *differentially* affect the types of financing of the firms that become more vs. less similar in the MSA with the arrivals of such shocks. Therefore, those external shocks rather help us exploit a meaningful variation in the changes in similarity scores across firms within the MSA. We also alleviate this concern further by examining VC funding instead. If a regional economic factor both attracts new startups (thus raising dissimilarity) and increase the local funding opportunities at the same time, we should observe the same directional effects with VC funding. Internet Appendix Table IA.5 report the results from the analogous tests using VC funding. We find strikingly opposite results from the table on VC funding. That is, a decrease in business similarity *decreases* firms' access to VC financing. These contrasting results for angel vs. VC financing are consistent with our previous analyses on the different investment styles of angel between VC investors in Table 3 and also effectively mitigate the concern about regional economic shocks.

It is also possible that our results are driven by the correlation between similarity and funded firm quality, which affects the demand for a certain type of early-stage financing. For example, the low similarity of firms in an MSA is associated with low competition in the region, therefore indicating a lack of high quality firms that would have been chosen to receive funding from VC otherwise. We rule out this possibility in Internet Appendix Table IA.6. We consider a regression model of the Entrepreneurship Quality Index (EQI) on MSA-level dissimilarity and find that high MSA-level dissimilarity rather improves the regional entrepreneurship quality, even after controlling for a number of MSA-level economic factors and entrepreneurial indices (entrepreneurship cohort potential index and regional ecosystem acceleration index). This result has two implications. First, our results are not driven by low quality firms left to receive angel funding because they were unable to receive VC funding. Second, given that EQI is based on the probability of observing growth outcome (*i.e.,* IPO success), the positive correlation between MSA-level dissimilarity and EQI is likely to convey fiercer entrepreneurial environments for survival when industrial agglomeration is high in a region.

To address the concern that angel funding data from Crunchbase may not be complete, we follow Denes et al. (2020) and supplement angel investment data with the SEC Form-D filings. We only use firms that raise equity and aggregate the total amount raised in each filing. We find that our results remain robust in Internet Appendix Table IA.7.

As the last but important analysis for robustness, we consider a broader spatial unit using Combined Statistical Area (CSA)[17]. This is to rule out the possibility that our results are predominantly driven by smaller MSAs, where angel investors may mechanically appear to provide more funding given their preference for geographically closer funded firms. For robustness to using a narrower spatial unit as well, we also consider counties instead of MSA units. We re-run all our analyses on business similarity and angel funding using both CSAs and counties and report the results in Internet Appendix Table IA.8. We find that our

[17]Combined Statistical Areas represent groupings of Metropolitan and Micropolitan Statistical Areas (in any combination) and can be characterized as representing larger regions that reflect broader social and economic interactions.

findings hold robust to using the broader or narrower spatial unit.

Collectively, all results in this section provide strong evidence that angel investors prefer firms that stand out from their local peers as unique businesses. This financing channel plays a significant role in startups' choosing a location to launch their business.

5.5 Angel Investor Diversity

In this section, we turn our focus to angel investors and examine whether specific characteristics of angel investors strengthen the relation between the business dissimilarity of startups and their angel funding. Angel investors rely on personal networks in the areas where they live to locate investment opportunities. More importantly, they often participate in local angel groups to improve their investment decisions by learning from other group members (Cable, 2011). Therefore, we are particularly interested in knowing how the local angel diversity affects their investment decisions to fund more diverse firms. We consider two dimensions of individual angel investors' diversity: Demography and education of local angel investors.

In Table 8, we repeat the regression analyses in Table 6 but additionally interact *Dissimilarity* with *High Diversity*, a MSA-year level indicator variable that equals to one if the diversity measure based on either demographics or education is above the median and zero otherwise. The diversity measure is computed as the sum of normalized diversity components by its mean and standard deviation where each diversity component is calculated as 1-(HHI of a component). Demographics consists of gender and race (Asian-pacific, Black, and White) components, and education consists of college and major components. Panels A and B report the results based on the MSA-level demographic and educational diversity of angel investors, respectively. The main variable of interest in this analysis is the interaction term between *Dissimilarity* with *High Diversity*.

Columns (1) and (2) in Panel A show when the region has more diverse angel investors in their demographic characteristics, startup firms with more diverse businesses have significantly higher likelihood of receiving angel funding and also the funding amount from angel investors are significantly higher. The incremental effect from the high diversity of angel investors is approximately 5% of the primary effect of the startup dissimilarity and statistically significant at the 1% level. We find that this finding is robust to using both the vintage and dynamic similarity measures alternatively. The results in Panel B based on the educational diversity are broadly consistent with those in Panel A and indicate that diverse startup firms are more likely to be funded by angel investors when those investors' educational backgrounds are more diverse.

The findings in this section are especially noteworthy in that investor diversity in a region is economically linked to local entrepreneurial financing. Our evidence has important policy implications that individual diversity in a region can be a driving force to nurture unique business ideas among startup firms.

6 Entrepreneurial Outcome

6.1 Investor Success

In this section, we investigate the funding outcomes of angel investors focusing on whether their preference for diverse-industry funding leads to more successful exits of their funded firms. Specifically, we examine whether an angel funding is followed by a subsequent funding round and the likelihood of a funded firm's exit through an IPO or acquisition. We estimate the following regression specification for this analysis:

$$Outcome_j = \alpha + \beta_1 Angel_j + \beta_2 Dissimilar_j + \beta_3 Angel_j \times Dissimilar_j + \gamma \Gamma_j + \epsilon_j, \qquad (5)$$

where $Outcome_j$ is one of the dependent variables that we consider including an indicator variable for a subsequent funding after investor j makes a funding, the average IPO exit rate among all funded firms by investor j, and the average successful exit rate including both IPO and acquisitions exits among all funded firms by investor j. Our main variable of interest is $Dissimilar_j$ which is an indicator variable if investor j's investment portfolio is identified to follow the diverse-industry strategy. An investor's portfolio is considered to follow the diverse-industry strategy if the overall similarity score among the funded firms in the investor's portfolio is below the median of all portfolios of investors of the same type. We then consider the interaction between $Dissimilar_j$ and the angel investor indicator for investor j ($Angel_j$) to examine whether the diverse-industry strategy only concerns angel investment strategies. Γ_j is a set of control variables at the investor level including the log number of total investments made by investor j, the log of investing years since investor j made her first investment, and the log of the average funding amount made by investor j. Table 9 presents the results.

[Insert Table 9 Here]

Columns (1) reports the regression results for subsequent investments. First, we note that the angel investor indicator is positive and significant in predicting subsequent funding rounds. This indicates that angel funding is more likely to come before any other funding type, which is discussed in the previous section regarding possible sequential investments between angel and VC. More importantly, we find that the interaction term of *Angel* with *Dissimilar* is significantly positive at the 1% level, implying that the diverse-industry strategy has a positive effect on the likelihood of an interim success of funded firms by an angel investor manifested as the incidence of the next funding rounds. The economic effect translates into a 13.6 percentage point increase in the likelihood that funded firms in the angel investor's portfolio receive the next rounds of funding. This effect is equivalent to a 20% increase from the unconditional mean of the subsequent-funding likelihood. In column (2), we report results from the analogous test by replacing the dependent variable with the fraction of firms with eventual IPO exits in the investor's portfolio. We first note that angel investors in general have significantly lower IPO exit rates relative to other types of investors. We also find that the diverse-industry strategy has a significantly positive effect on the IPO exit rate of an angel investor's portfolio. The estimated effect is a 3.4 percentage point increase in the portfolio's IPO exit rate. We find stronger results by additionally considering a funded firm's being acquired as a successful outcome in column (3). Angel investors' successful exit rates, both in IPO and acquisition exits, increase by 7.0 percentage points or 32% from the unconditional mean of exit rates by following the diverse-industry strategy.[18]

Overall, these results are consistent with the interpretation that angel investors by diversifying their portfolios over the dimension of industry classifications (with maintaining close geographical proximity to funded firms) can optimally select the net-present-value positive investments from their investment opportunity sets. We find that angel investors can significantly improve their successful exit rates without forgoing optimal investment opportunities with proper diversification.[19]

6.2 Firm Success

We now turn to entrepreneurial outcomes from the perspective of firms. We expect that firms that are disparate in a region successfully solicit angel investors' attention and investment and thus are more likely to be successful eventually. If this was not the case, we would not observe the evidence of startup firms choosing to locate in MSAs with a greater number of angel investors. We thus examine the relation between a firm's similarity score relative to its geographical peers and measures of firm success. We specifically estimate the following regression specification for this analysis:

$$\begin{aligned} \mathit{Firm\ Outcome}_j &= \alpha + \beta_1 \mathit{Initial\ Similarity}_j + \beta_2 \mathbb{1}(\mathit{Angel\ Funding}) \\ &\quad + \beta_3 \mathit{Initial\ Similarity}_j \times \mathbb{1}(\mathit{Angel\ Funding}) + \gamma \Gamma_j + \epsilon_j, \end{aligned} \tag{6}$$

[18]There may be a concern that investors with only one investment may drive our results. However, our similarity score cannot be computed when there is only one investment, and thus the case falls into $\mathit{Dissimilar} = 0$. We also find consistent results after excluding investors with only one investment.

[19]Consistent with this interpretation, the average number of investments is significantly greater at 7.34 for angel investors with the diverse-industry strategy, relative to 2.03 for angel investors with no such strategy.

where $Firm\ Outcome_j$ is one of the dependent variables that we consider including firm j's total number of funding rounds, total amount of funding received, and likelihood of an exit through an IPO or acquisition. Our main variable of interest is the interaction term between $Initial\ Similarity_j$ and $\mathbb{1}(Angel\ Funding)$. $\mathbb{1}(Angel\ Funding)$ is an indicator variable if firm j has ever received angel funding during our sample period.[20] Γ_j is a set of control variables at the firm j's MSA level including the log of the total number of firms and E and EG indexes. This test is cross-sectional across firms in our sample, and thus the sample consists of one observation per firm. Table 10 presents the results.

[**Insert Table 10 Here**]

Columns (1) and (2) report the regression results for the successful funding outcomes. We first find that whether a firm has any angel funding round or not is a good indicator for its total number of funding rounds and total funding amount throughout the sample period. The economic effect is that a firm with any angel financing has 1.6 more funding rounds in total and also $1.7 million more in the funding amount. When we interact the indicator for an angel funding with a firm's similarity score at its entry in both columns, we find that the positive effect of angel financing on total funding outcomes strengthens with the increased dissimilarity. For example in column (2), when a firm entered into a market with a similarity score lower than others by one standard deviation and then received angel financing, its total funding amount would be approximately $1.1 million higher. In contrast, the interaction term between the indicator for VC financing and the initial similarity score is significantly positive, indicating that for a firm that entered into a market with a higher similarity score, receiving VC financing increases the total funding amount significantly.

In columns (3) and (4), we report results from the analogous tests by replacing the dependent variable with indicators for an IPO exit and either an IPO or acquisition exit,

[20] For this analysis, we use the initial similarity score at a firm's entry. Because any financing round is always after a firm's entry, using the initial similarity score helps mitigate commingled effects between changing similarity scores and the likelihood of angel financing over time. Our results are robust to using the average similarity score over the sample period.

respectively. Although the indicator for any angel financing has no effect on the likelihood of an IPO exit, it appears to significantly increase the likelihood of the combined exit of an IPO or acquisition as shown in column (4). The coefficient estimates for the interaction term between the indicator for angel financing and the initial similarity score in both columns continue to be negative and significant, albeit weaker in the statistical significance at the 10% level. For example in column (4), the economic interpretation of the effect is that when a firm entered into a market with a similarity score lower than others by one standard deviation and then received angel financing, its likelihood of an IPO or acquisition exit would increase by 0.98 percentage points or 10% from the unconditional mean. On the other hand, we find the opposite results for VC financing that for a firm that entered into a market with a higher similarity score, receiving VC financing increases its chance to exit with an IPO or acquisition.

Overall, these results on firm success portray two possible success strategies for new firm formation. One is to target angel financing after entering into a region where the firm can stand out as a disparate business. The other is to target VC financing after entering into a region where similar firms have already formed industry clustering. We find that while the latter strategy of new firm formation has been prevailing in the past, the former strategy emerges more recently leading to the geographic dispersion of new firms and industry diversity of the regions where they enter.

7 Conclusion

In this paper, we introduce a new measure for startup business similarity using detailed industry classifications available from the Crunchbase data. This measure allows us to examine time-varying and across-region firm similarity and its effects on the types of early-stage financing. We find that increasingly more startup firms enter a regional market when their local peers are more dissimilar to them. This unique trend appears to be associated with local angel investment opportunities as our evidence shows that dissimilar startup firms are more likely to be funded by angel investors.

We also document that individual investors' demographic and educational diversity in a region plays an important role to reinforce the relation between the business diversity and angel financing. Our results suggest that policies designed to increase the number of angel investors and their demographic or educational diversity may nurture unique business ideas. Finally, we show that successful exit rates of angel investors' portfolios significantly improve when they fund diverse-industry startup firms. Collectively, our results offer new insight into the specific mechanism for the new trend of firm formation, which is geographic dispersion with industry diversity.

References

Andrews, Raymond, Catherine Fazio, Jorge Guzman, Yupeng Liu, and Scott Stern, 2019, The startup cartography project: Measuring and mapping entrepreneurial ecosystems Working Paper.

Arrow, Kenneth Joseph, 1962, The economic implications of learning by doing, *Review of Economic Studies, 29 (1962), pp. 155-172* 29, 155–172.

Beaudry, Catherine, and Andrea Schiffauerova, 2009, Who's right, marshall or jacobs? the localization versus urbanization debate, *Research policy* 38, 318–337.

Bernstein, Shai, Arthur Korteweg, and Kevin Laws, 2017, Attracting early-stage investors: Evidence from a randomized field experiment, *Journal of Finance* 62, 509–538.

Cable, Abraham, 2011, Fending for themselves: Why securities regulations should encourage angel groups, *U. of Penssylvania Journal of Business Law* 13, 107–172.

Carlino, Gerald, and William Kerr, 2015, Agglomeration and innovation 5, edited by Gilles Duranton, J. Vernon Henderson, and William Strange, 349–404, Handbook of Regional and Urban Economics.

Chatterji, Aaron, Edward Glaeser, and William Kerr, 2014, Clusters of entrepreneurship and innovation 14, In: Lerner, J., Stern, S. (Eds.), Innovation Policy and the Economy.

Chemmanur, Thomas J, and Zhaohui Chen, 2014, Venture capitalists versus angels: The dynamics of private firm financing contracts, *The Review of Corporate Finance Studies* 3, 39–86.

Chen, Henry, Paul Gompers, Anna Kovner, and Josh Lerner, 2009, Buy local? the geography of successful and unsuccessful venture capital expansion Working paper.

Chen, Henry, Paul Gompers, Anna Kovner, and Josh Lerner, 2010, Buy local? the geography of successful and unsuccessful venture capital expansion, *Journal of Urban Economics* 67, 90–102.

Chen, Jun, and Michael Ewens, 2021, Venture capitalists' access to finance and its impact on startups, *NBER Working Paper* .

Davis, Jesse, Adair Morse, and Xinxin Wang, 2020, The leveraging of silicon valley Working Paper.

Denes, Matthew, Sabrina Howell, Filippo Mezzanotti, Xinxin Wang, and Ting Xu, 2020, Investor tax credits and entrepreneurship: Evidence from us states Working Paper, Carnegie Mellon University.

Ellison, Glenn, and Edward Glaeser, 1997, Geographic concentration in u.s. manufacturing industries: A dartboard approach, *Journal of Political Economy* 105, 889–927.

Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf, 2018, Cost of experimentation and the evolution of venture capital, *Journal of Financial Economics* 128, 422–442.

Fulghieri, Paolo, and Merih Sevilir, 2009, Size and focus of a venture capitalist's portfolio, *The Review of Financial Studies* 22, 4643–4680.

Glaeser, Edward, William Kerr, and Giacomo Ponzetto, 2010, Clusters of entrepreneurship, *Journal of Urban Economics* 67, 150–168.

Gompers, Paul, 1995, Optimal investment, monitoring, and the staging of venture capital, *Journal of Finance* 50, 1461–1489.

Gompers, Paul, Will Gornall, Steven Kaplan, and Ilya Strebulaev, 2020, How do venture capitalists make decisions?, *Journal of Financial Economics* 135, 169–190.

Gornall, Will, and Ilya Strebulaev, 2021, The economic impact of venture capital: Evidence form public companies, *Working paper* .

Hellmann, Thomas, and Veikko Thiele, 2015, Friends or foes? the interrelationship between angel and venture capital markets, *Journal of Financial Economics* 115, 639–653.

Hellmann, Thomas F, Paul Schure, and Dan Vo, 2021, Angels and venture capitalists: substitutes or complements?, *Journal of Financial Economics* forthcoming.

Higgins, Keith, 2014, The changing regulatory landscape for angel investing. Harvard Law School Forum on Corporate Governance. https://corpgov.law.harvard.edu/2014/04/21/the-changing-regulatory-landscape-for-angel-investing/ .

Hochberg, Yael, Alexander Ljungqvist, and Yang Lu, 2007, Whom you know matters: Venture capital networks and investment performance, *Journal of Finances* 62, 251–302.

Hochberg, Yael, Alexander Ljungqvist, and Yang Lu, 2010, Networking as a barrier to entry and the competitive supply of venture capital, *Journal of Finances* 65, 829–859.

Hochberg, Yael, Michael Mazzeo, and Ryan McDevitt, 2015, Specialization and competition in the venture capital industry, *Review of Industrial Organization* 46, 323–347.

Jacobs, Jane, 1969, Economy of cities, *New York: Vintage* .

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, 1993, Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* 108, 577–598.

Kaplan, Steven, and Josh Lerner, 2016, Venture capital data: Opportunities and challenges NBER Working Paper.

Kaplan, Steven, and Per Stromberg, 2001, Venture capitalists as principals: Contracting, screening, and monitoring, *American Economic Review* 91, 426–430.

Kauffman Foundation, 2002, Business angel investing groups growing in north america Kauffman Foundation.

Kerr, William, Josh Lerner, and Antoinette Schoar, 2014, The consequences of entrepreneurial finance: Evidence from angel financing, *Review of Financial Studies* 27, 20–55.

Kolympiris, Christos, Nicholas Kalaitzandonakes, and Douglas Miller, 2011, Spatial collocation and venture capital in the us biotechnology industry, *Research Policy* 40, 1188–1199.

Lerner, Josh, Antoinette Schoar, Stanislav Sokolinski, and Karen Wilson, 2018, The globalization of angel investments: Evidence across countries, *Journal of Financial Economics* 127, 1–20.

Lerner, Joshua, 1995, Venture capitalists and the oversight of private firms, *Journal of Finance* 50, 301–318.

Lerner, Joshua, 2020, Venture capital's role in financing innovation: What we know and how much we still need to learn, *Journal of Economic Perspectives* 34, 237–261.

Lindsey, Laura, and Luke Stein, 2019, Angels, entrepreneurship, and employment dynamics: Evidence from investor accreditation rules. Working paper.

Marshall, Alfred, 1890, Principles of economics, *London Mcmillan* .

Marshall, Alfred, 1920, *Principles of Economics* (MacMillan).

MIT Entrepreneurship Center, 2000, Venture support systems project: Angel investors MIT Entrepreneurship Center.

Puri, Manju, and Rebecca Zarutskie, 2012, On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms, *The Journal of Finance* 67, 2247–2293.

Romer, Paul M, 1986, Increasing returns and long-run growth, *Journal of political economy* 94, 1002–1037.

Rosenthal, Stuart, and William Strange, 2008, The attenuation of human capital spillovers, *Journal of Urban Economics* 64, 373–389.

Sidorov, Grigori, Alexander Gelbukh, Helena Gómez-Adorno, and David Pinto, 2014, Soft similarity and soft cosine measure: Similarity of features in vector space model, *Computación y Sistemas* 18, 491–504.

Sorenson, Olav, and Toby Stuart, 2001, Syndication networks and the spatial distribution of venture capital investments, *American Journal of Sociology* 106, 1546–1588.

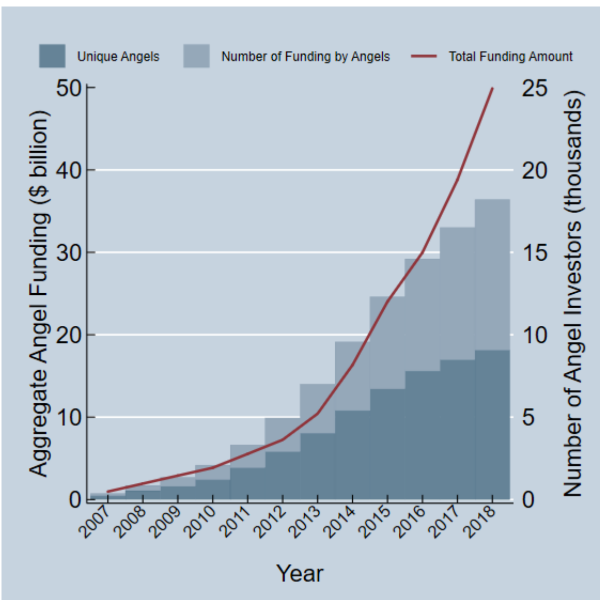
Wang, Xinxin, 2018, Catering innovation: Entrepreneurship and the acquisition market Working Paper.

Wharton Entrepreneurship and Angel Capital Association, 2017, The american angel Wharton Entrepreneurship and Angel Capital Association.

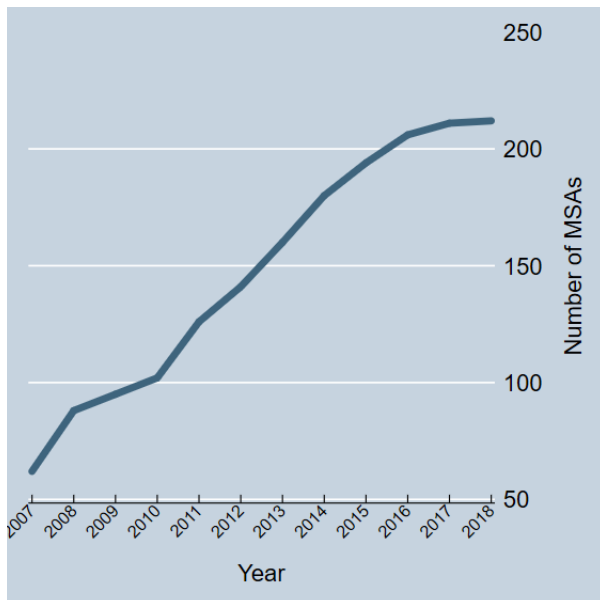
Wong, Andrew, 2002, Angel finance: The other venture capital Working paper.

Figure 1: Growth of Angel Investors and Angel Funding Over Time

The figures show the cumulative angel funding amount and the number of angel investors in Crunchbase over our sample period 2007-2018. In Panel (a), the darker blue bar represents the cumulative number of unique angel investors, and the lighter blue bar represents the cumulative number of funding financed by our sample angel investors. The red line shows the cumulative total dollar amount of angel funding in $ billion. In Panel (b), we show the number of MSAs with at least one Crunchbase angel investor over time. The total number of MSAs is 416.



(a) Cumulative aggregate funding amount and number of angels



(b) Number of MSAs with angel investor presence

Figure 2: Heterogeneity in Portfolio Strategy by Angel Investment Size

The figure shows the angel investor portfolio strategy by investment size. *Industry diversity* measures the business similarity among an angel's portfolio firms. *Geographic distance* measures the average distance between an angel and its portfolio firms. The colors of the bubbles denote the size deciles of angel investment, where the color becomes closer to blue as the investment size grows. Each bubble represents the average geographic distance and industry diversity of the given angel investments in a given size decile. For variable definitions and further details of their construction, see Internet Appendix C.

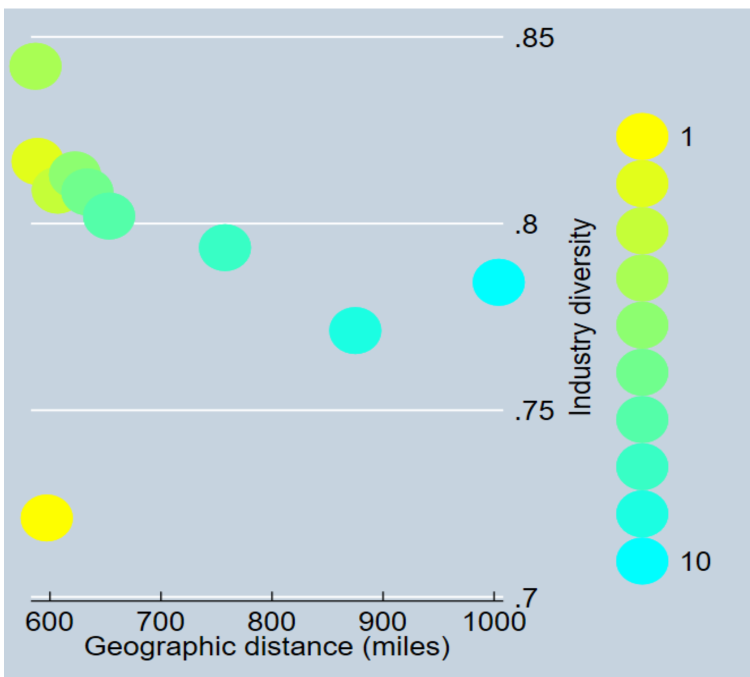


Figure 3: Post-entry Variations of Similarity Scores

The figure shows variations of similarity scores over firm age after they enter an MSA in the form of box plots. Each box displays the interquartile range between the 25th to 75th percentiles of the distribution of the changes in similarity, where the solid line inside the box represents the median. The top and bottom solid lines outside the box display the maximum and minimum, respectively, where the maximum and minimum are defined as the 75th percentile+1.5×the interquartile range and the 25th percentile−1.5×the interquartile range.

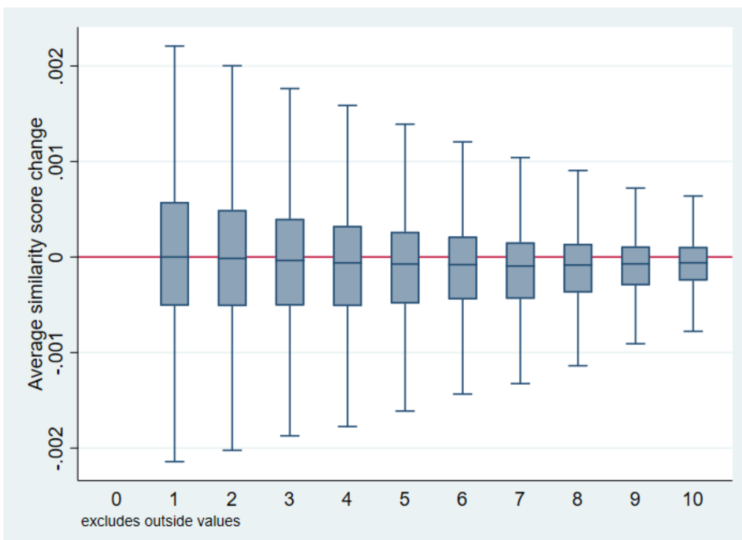


Table 1: Rise of New Startup Hubs

The table shows the rising startup hubs and angel investor growths. Panel A lists the top 25 (excluding San Francisco and San Jose) rising startup hubs (MSAs) by the number of firm growth over the 2007-2018 sample period. The MSAs considered Silicon Valley, which are San Francisco and San Jose, are denoted with *. MSAs that are often mentioned as new rising startup hubs in media are denoted with †, which come from media sources such as Bloomberg, the Economist, and Kenan Insights report as referenced in the footnote 2. Panel B shows the differences in means of the angel growth rates and the p-values. The upper panel divides the sample into above (High) and below (Low) the median startup growth rate MSAs, and the lower panel divides the sample into top 25 (Panel A) and other MSAs by the startup growth rates.

Panel A: Top 25 growing hubs

MSA Name	Firm growth (%)
*San Francisco-Oakland-Fremont, CA	138.3
Las Vegas-Paradise, NV	121.1
Provo-Orem, UT	110.1
†Austin-Round Rock, TX	108.5
†New York-Northern New Jersey-Long Island, NY-NJ-PA	102.1
†Boulder, CO	102.1
*San Jose-Sunnyvale-Santa Clara, CA	100.4
Miami-Fort Lauderdale-Pompano Beach, FL	93.8
†Los Angeles-Long Beach-Santa Ana, CA	93.0
Charleston-North Charleston, SC	88.5
†Seattle-Tacoma-Bellevue, WA	87.0
Boise City-Nampa, ID	86.9
†Durham, NC	85.3
†Denver-Aurora, CO	79.2
†San Diego-Carlsbad-San Marcos, CA	78.6
New Orleans-Metairie-Kenner, LA	78.0
Ann Arbor, MI	77.5
Nashville-Davidson-Murfreesboro-Franklin, TN	75.6
Riverside-San Bernardino-Ontario, CA	72.4
Reno-Sparks, NV	72.2
†Boston-Cambridge-Quincy, MA-NH	70.3
†Columbus, OH	70.2
Orlando-Kissimmee, FL	70.2
†Chicago-Naperville-Joliet, IL-IN-WI	69.1
Santa Cruz-Watsonville, CA	68.0
Sarasota-Bradenton-Venice, FL	68.0
†Portland-Vancouver-Beaverton, OR-WA	67.9

Panel B: Growing angel investors

	Avg. angel growth rate (%)			
Period	High startup growth	Low startup growth	Diff	p-value
2007-2018	22.85	15.46	7.38	0.0002
2007-2012	24.09	15.70	8.39	0.0360
2013-2018	22.17	15.38	6.79	0.0022

	Avg. angel growth rate (%)			
Period	Top 25	Other	Diff	p-value
2007-2018	28.53	18.38	10.15	0.0000
2007-2012	30.63	18.80	11.83	0.0066
2013-2018	26.95	18.20	8,75	0.0047

Table 2: Summary Statistics of Crunchbase Variables

The full sample for the analyses consists of 119,605 unique firms between 2007-2018. The top panel presents firm-level summary statistics on the funding sample, which consists of 18,451 firms that report at least one funding round between 2007-2018 (except for *All firm age*). The middle and bottom panels present descriptive statistics on funding by investor types using the funding sample. There are 47,121 funding round-investor type observations. *Fraction of funding* is the fraction of the funding amount provided by the investor type on a given funding round. *Distance to investors* is the distance measured in miles between the funded firm and the given type of funding investor. For variable definitions and further details of their construction, see Internet Appendix C.

	mean	sd	min	p50	max	N
Firm-level						
All firm age (snapshot in 2018)	5.51	2.73	0	6	10	73,822
Funded firm age (snapshot in 2018)	5.53	2.50	0	5	10	13,087
Number of funding rounds	1.90	1.33	1	1	29	18,451
Age at funding	3.07	2.49	0	3	10	18,451
Age at first funding	2.48	2.48	0	2	10	18,451
Funding-level: Angel						
Angel (dummy)	0.19	0.39	0	0	1	47,121
Funded firm age	2.35	2.13	0	2	10	9,052
Funded firm age at first funding	1.75	1.91	0	1	10	5,828
Fraction of first funding	0.22	0.41	0	0	1	26,469
Number of investors	2.07	2.14	1	1	61	9,052
Funding amount ('000s)	2,427.93	9,469.87	0	875	499,505	9,052
Fraction of funding	0.43	0.33	0	0	1	9,052
Distance to investors (miles)	549.79	791.05	0	88	3,294	8,774
Funding-level: VC						
VC (dummy)	0.59	0.49	0	1	1	47,121
Funded firm age	3.37	2.59	0	3	10	27,694
Funded firm age at first funding	2.53	2.46	0	2	10	14,742
Fraction of first funding	0.56	0.50	0	1	1	26,469
Number of investors	2.41	1.74	1	2	31	27,694
Funding amount ('000s)	11,459.04	45,980.61	1	4,286	4,620,000	27,694
Fraction of funding	0.52	0.33	0	1	1	27,694
Distance to investors (miles)	724.35	809.61	0	427	4,972	27,428

Table 3: Angel vs. VC Portfolio Preferences

The table shows results from the regressions of investor portfolio characteristics on investor type. Regression in Panels A and B are at the aggregate investor level and each funding level, respectively. We consider within-investor industry category similarities in columns (1) and (2) and the geographic distance in columns (3) and (4). For those measures, we consider all funding during the entire sample period by each investor in Panel A and previous funding by each investor before making a given funding in a given year in Panel B. The sample in columns (1) and (3) consists of all investor types, and the sample in columns (2) and (4) consists of only VC and angel investors. $Angel$ and VC are indicator variables. $\$Funding$ is the average funding amount, $Ln(Investments)$ is the log of the total number of investments, and $Ln(InvestingYears)$ is the log of the years since the first investment. $Ln(Investments)$ and $\$Funding$ in Panel B only consider previous investments before a given funding year. Standard errors are clustered at the investor MSA level. For variable definitions and further details of their construction, see Internet Appendix C.

Panel A: Investor-level

	Industry Similarity		Geographic Distance	
	(1)	(2)	(3)	(4)
Angel	0.009	-0.018*	-93.914**	-106.908***
	(0.011)	(0.011)	(42.571)	(32.192)
VC	0.031***		20.988	
	(0.011)		(25.017)	
Ln(Investments)	-0.054***	-0.058***	-7.638	-8.484
	(0.005)	(0.006)	(7.573)	(7.351)
Ln(Investing Years)	-0.019	-0.055***	-79.014	-77.079
	(0.018)	(0.016)	(54.991)	(67.447)
$Funding	0.000	0.001**	4.360***	5.020***
	(0.000)	(0.000)	(0.571)	(1.196)
Observations	7,185	6,070	11,853	10,384
Adjusted R^2	0.076	0.081	0.068	0.060
Investor MSA FE	Y	Y	Y	Y

Panel B: Funding-level

	Industry Similarity		Geographic Distance	
	(1)	(2)	(3)	(4)
Angel	-0.018***	-0.018***	-154.580***	-52.453***
	(0.005)	(0.004)	(33.119)	(15.073)
VC	0.002		-108.195***	
	(0.004)		(27.449)	
Ln(Investments)	-0.012***	-0.013***	25.426***	26.962***
	(0.002)	(0.001)	(8.354)	(7.934)
Ln(Investing Years)	0.013***	0.010***	79.431***	68.528***
	(0.005)	(0.002)	(16.851)	(19.850)
$Funding	0.000	0.000**	3.535***	2.264**
	(0.000)	(0.000)	(0.774)	(0.989)
Observations	45,158	39,326	44,445	38,732
Adjusted R^2	0.571	0.582	0.251	0.255
Firm FE	Y	Y	Y	Y
Funding year FE	Y	Y	Y	Y
Investor MSA FE	Y	Y	Y	Y
Funded MSA FE	Y	Y	Y	Y

Standard are errors in parentheses.
***p<0.01, **p<0.05, *p<0.1

Table 4: Angel Investors and Local Startup Similarity

The table presents regressions of local similarity at on a lagged angel investor presence. *Similarity* is the similarity scores. Column (1) uses the enactment of Dodd-Frank Act as an exogenous shock to the number of accredited investors. The sample in column (1) only includes the pre-period of the General Solicitation rule change in 2013. *Treated* in column (1) is an indicator variable, which is equal to one if the fraction of families that may have lost accreditation due to the act at the state level is *below* median and zero otherwise. *Post* in column (1) is an indicator variable, which is equal to 1 if after the enactment of Dodd-Frank in 2010 and 0 otherwise. Column (2) uses the elimination of general solicitation prohibition of Rule 506(c) as an exogenous shock to the number of accredited investors. The sample in column (2) only includes the post-Dodd-Frank period after 2010. *Treated* in column (2) is an indicator variable, which equals to 1 if the number of accredited investors at the family level during the pre-elimination period is greater than the median and 0 otherwise. *Post* in column (2) is an indicator variable, which is equal to 1 if after the elimination in 2013 and 0 otherwise. All columns control for the one-year lagged local economic variables. Standard errors are clustered at the MSA-year level. For variable definitions and further details of their construction, see Internet Appendix C.

	Similarity × 100	
	Dodd-Frank	General Solicitation
	(1)	(2)
Treated×post	-0.1500**	-0.1600**
	(0.068)	(0.071)
Ln(Firms)	-1.2282**	0.5828
	(0.514)	(0.693)
Ln(Firm Deaths)	-0.2301	0.3437
	(0.282)	(0.321)
Ln(Investors)	-0.0172	-0.1529*
	(0.109)	(0.084)
Ln(Population)	5.5765	-0.1263
	(3.718)	(1.186)
Ln(GDP)	-2.0848***	-0.1842
	(0.629)	(0.468)
Observations	1090	1704
Adjusted R^2	0.850	0.811
MSA FE	Y	Y
Year FE	Y	Y

Standard errors are in parentheses.
***p<0.01, **p<0.05, *p<0.1

Table 5: Angel Investors and Startup Entry Decision

The table presents cross-sectional regressions of entry decisions on similarity and angel investor presence. The sample is firm-MSA-entry year level observations for the 81,718 entering firms between 2007-2018. *Entry* is an indicator variable equal to 1 if a given MSA is an ultimate entering location and 0 otherwise. *Similarity* is the similarity scores. Each entering firm gets 412 similarity scores–one for each potential entering MSA using the existing firms in a given MSA in the year prior to the entry. Column (1) uses the enactment of Dodd-Frank Act as an exogenous shock to the number of accredited investors. The sample in column (1) only includes the pre-period of the General Solicitation rule change in 2013. *Treated* in column (1) is a continuous variable, which is computed as taking the negative value of the fraction of families that may have lost accreditation due to the act at the state level. *Post* in column (1) is an indicator variable, which is equal to 1 if after the enactment of Dodd-Frank in 2010 and 0 otherwise. Column (2) uses the elimination of general solicitation prohibition of Rule 506(c) as an exogenous shock to the number of accredited investors. *Treated* in column (2) is a continuous variable, which measures the number of accredited investors at the family level. *Post* in column (2) is an indicator variable, which is equal to 1 if after the elimination in 2013 and 0 otherwise. The sample in column (2) only includes the post-Dodd-Frank period after 2010. All columns control for the local economic variables, and column (2) additionally controls for the quality-adjusted quantity of entrepreneurship using the Regional Ecosystem Acceleration Index (REAI) from (Andrews et al., 2019), which ends in 2013 and thus is not included in column (2). The control variables are measured in the year prior to the entry year. Standard errors are clustered at the firm level. For variable definitions and further details of their construction, see Internet Appendix C.

	Entry × 100	
	Dodd-Frank	General Solicitation
	(1)	(2)
Dissimilarity×treated×post	2.6398**	0.0126***
	(1.093)	(0.003)
Treated×post	-2.4738**	-0.0125***
	(1.067)	(0.003)
Dissimilarity	1.2463***	1.6701***
	(0.279)	(0.135)
Ln(Investors)	0.7135***	0.6299***
	(0.004)	(0.003)
Ln(Firms)	-0.2139***	-0.2122***
	(0.006)	(0.005)
Ln(Personal Income)	0.6501***	0.8107***
	(0.018)	(0.020)
Ln(GDP)	0.1018***	0.1277***
	(0.007)	(0.006)
REAI	-0.0240***	
	(0.001)	
Observations	10312898	9540949
Adjusted R^2	0.024	0.025
MSA FE	Y	Y
Entry year FE	Y	Y

Standard errors are in parentheses.
***p<0.01, **p<0.05, *p<0.1

Table 6: Startup Similarity and Angel Funding

The table shows results from the firm-year level regressions of funding characteristics on dissimilarity. The main variable of interest is *Dissimilarity*. The sample replaces no-funding observations with zeros. $\mathbb{1}(Angel)$ is a dummy variable that is equal to one if funding is received from angel investors and zero otherwise. $\$Funding$ is the total amount of funding in millions from angel investors in the given year. All specifications control for G and EG indices. Standard errors are clustered at the MSA level. For variable definitions and further details of their construction, see Internet Appendix C.

	MSA		Vintage		Soft-cosine		Dynamic	
	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dissimilarity	0.359***	0.149***	0.371***	0.159***	0.214***	0.079***	0.133**	0.077**
	(0.129)	(0.054)	(0.134)	(0.058)	(0.041)	(0.019)	(0.064)	(0.035)
Ln(Age)	0.002	-0.001*	0.002	-0.001*	0.002	-0.001	0.001	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(Firms)	-0.057***	-0.026***	-0.058***	-0.026***	-0.033***	-0.015***	-0.058***	-0.026***
	(0.013)	(0.005)	(0.014)	(0.005)	(0.010)	(0.004)	(0.014)	(0.005)
Observations	762299	762299	759230	759230	697342	697342	762187	762187
Adjusted R^2	0.123	0.093	0.123	0.093	0.123	0.070	0.122	0.093
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y

Table 7: Startup Similarity and Angel Funding – Robustness

The table shows the results from the firm-year level regressions of funding characteristics on similarity. Panel A drops the entry-year observations to mitigate selection concerns. Panel B restricts the sample to first funding observations only. We cannot include investor-level fixed effects when the first funding is funded by more than one type of investors. $\mathbb{1}(Angel)$ is a dummy variable that is equal to one if funding is received from angel investors and zero otherwise. *$Funding* is the total amount of funding in millions from angel investors in the given year. All specifications control for G and EG indices. Standard errors are clustered at the MSA level in Panel A and funded firm MSA level in Panel B. For variable definitions and further details of their construction, see Internet Appendix C.

Panel A: Dropping entry-year observations

	MSA		Vintage		Soft-cosine		Dynamic	
	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)		
Dissimilarity	0.381***	0.176***	0.396***	0.188***	0.224***	0.098***	0.132**	0.078**
	(0.138)	(0.058)	(0.145)	(0.063)	(0.040)	(0.022)	(0.065)	(0.034)
Ln(Age)	-0.022***	-0.015***	-0.022***	-0.015***	-0.015***	-0.011***	-0.022***	-0.015***
	(0.007)	(0.005)	(0.007)	(0.005)	(0.004)	(0.002)	(0.007)	(0.005)
Ln(Firms)	-0.062***	-0.025***	-0.062***	-0.026***	-0.034***	-0.014***	-0.062***	-0.026***
	(0.016)	(0.006)	(0.016)	(0.006)	(0.010)	(0.003)	(0.015)	(0.006)
Observations	678838	678838	676124	676124	621269	621269	678729	678729
Adjusted R^2	0.141	0.093	0.141	0.093	0.141	0.089	0.141	0.093
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: First-funding observations

	$\mathbb{1}$(Angel)			
	MSA	Vintage	Soft-cosine	Dynamic
	(1)	(2)	(3)	(4)
Dissimilarity	0.403***	0.429***	0.091	0.340***
	(0.082)	(0.077)	(0.065)	(0.102)
Ln(Firms)	-0.029	-0.030	-0.017	-0.032
	(0.040)	(0.040)	(0.101)	(0.040)
Observations	15349	15326	13188	15346
Adjusted R^2	0.042	0.042	0.046	0.041
Funding year FE	Y	Y	Y	Y
Founding year FE	Y	Y	Y	Y
Firm MSA FE	Y	Y	Y	Y
Investor MSA FE	Y	Y	Y	Y

Standard errors are in parentheses.
***p<0.01, **p<0.05, *p<0.1

Table 8: Angel Investor Diversity and Angel Funding

The table shows results from the firm-year level regressions of funding characteristics on similarity (Table 6) interacting with the diversity of local angel investors. The sample uses all sample observations by replacing no-funding observations with zeros. The main variable of interest is *Dissimilarity*. *High diversity* is a MSA-year level indicator variable equal to 1 if the diversity measure (*Education* or *Demographics*) is above the median and 0 otherwise. Each diversity measure is computed as the sum of normalized diversity components by its mean and standard deviation, where each diversity component is calculated as 1-(HHI of a component). *Education* consists of college and major components, and *Demographics* consists of gender and race (Asian-pacific, Black, and White) components. $\mathbb{1}(Angel)$ is a dummy variable that is equal to one if funding is received from angel investors and zero otherwise. $\$Funding$ is the total amount of funding in millions from angel investors in the given year. All specifications control for G and EG indices. Standard errors are clustered at the MSA level. For variable definitions and further details of their construction, see Internet Appendix C.

Panel A: Demographic diversity

	MSA		Vintage		Soft-cosine		Dynamic	
	$\mathbb{1}$(Angel)	\$Funding	$\mathbb{1}$(Angel)	\$Funding	$\mathbb{1}$(Angel)	\$Funding	$\mathbb{1}$(Angel)	\$Funding
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High diversity×Dissimilarity	0.038***	0.022***	0.036***	0.020***	-0.009	-0.005	0.023**	0.017**
	(0.013)	(0.007)	(0.013)	(0.008)	(0.015)	(0.009)	(0.010)	(0.007)
Dissimilarity	0.654***	0.294***	0.662***	0.306***	0.352***	0.147***	0.151**	0.083**
	(0.180)	(0.079)	(0.179)	(0.081)	(0.039)	(0.026)	(0.069)	(0.039)
High diversity	-0.036***	-0.021***	-0.034**	-0.019**	0.008	0.004	-0.022**	-0.016**
	(0.013)	(0.007)	(0.013)	(0.008)	(0.014)	(0.008)	(0.010)	(0.007)
Ln(Age)	0.001	-0.001	0.002	-0.001	0.002	-0.001	0.001	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(Firms)	-0.074***	-0.034***	-0.075***	-0.034***	-0.046***	-0.022***	-0.075***	-0.034***
	(0.013)	(0.005)	(0.013)	(0.005)	(0.012)	(0.004)	(0.013)	(0.005)
Observations	691296	691296	688227	688227	626339	626339	691176	691181
Adjusted R^2	0.121	0.090	0.121	0.090	0.121	0.066	0.121	0.090
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Education diversity

	MSA		Vintage		Soft-cosine		Dynamic	
	$\mathbb{1}$(Angel)	\$Funding	$\mathbb{1}$(Angel)	\$Funding	$\mathbb{1}$(Angel)	\$Funding	$\mathbb{1}$(Angel)	\$Funding
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High diversity×Dissimilarity	0.042***	0.009	0.042***	0.009	0.015**	0.003	0.043**	0.005
	(0.015)	(0.015)	(0.015)	(0.015)	(0.007)	(0.004)	(0.018)	(0.011)
Dissimilarity	0.731***	0.297***	0.734***	0.310***	0.366***	0.135***	0.158**	0.074*
	(0.187)	(0.084)	(0.183)	(0.086)	(0.043)	(0.026)	(0.070)	(0.038)
High diversity	-0.039***	-0.008	-0.039***	-0.008	-0.011*	-0.002	-0.040**	-0.004
	(0.014)	(0.014)	(0.014)	(0.014)	(0.007)	(0.003)	(0.017)	(0.011)
Ln(Age)	0.001	-0.002*	0.001	-0.002*	0.001	-0.001	0.001	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(Firms)	-0.080***	-0.036***	-0.082***	-0.036***	-0.052***	-0.024***	-0.081***	-0.036***
	(0.012)	(0.005)	(0.012)	(0.005)	(0.013)	(0.005)	(0.012)	(0.005)
Observations	641906	641906	638837	638837	576949	576949	641787	641787
Adjusted R^2	0.121	0.089	0.121	0.089	0.121	0.065	0.120	0.089
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Outcome of Investments

The table shows results from the regressions of successful outcomes on investor portfolio characteristics. Regressions are at the aggregate investor level with one observation per investor. We consider within-investor industry similarities and the geographic distance as portfolio characteristics. For those measures, we consider all fundings made by each investor during the entire sample period. *Dissimilar* which is an indicator variable if the investor's investment portfolio is identified to follow the diverse-industry strategy. An investor's portfolio is considered to follow the diverse-industry strategy if the overall similarity score among the funded firms in the investor's portfolio is below the median of all portfolios of investors of the same type. *Angel* is an indicator variable for angel investors. $Ln(Investments)$ is the log of the total number of investments by a given investor, $Ln(Investing\ Years)$ is the log of the years since the first investment of a given investor, and $\$Funding$ is the average funding amount made by a given investor. Standard errors are clustered at the investor MSA level. For variable definitions and further details of their construction, see Internet Appendix C.

	$\mathbb{1}$(Subsequent Funding)	Exit Rate - IPO	Exit Rate - IPO/Acq.
	(1)	(2)	(3)
Angel	0.069***	-0.050***	-0.034***
	(0.008)	(0.007)	(0.010)
Dissimilar	-0.075***	-0.043***	-0.061***
	(0.009)	(0.005)	(0.008)
Dissimilar × Angel	0.136***	0.034***	0.070***
	(0.012)	(0.005)	(0.011)
Ln(Investments)	0.174***	0.001	-0.006*
	(0.009)	(0.002)	(0.003)
Ln(Investing Years)	-0.120***	-0.118***	-0.677***
	(0.041)	(0.011)	(0.029)
$Funding	-0.000**	0.000***	0.000**
	(0.000)	(0.000)	(0.000)
Mean of the Dependent Variable	0.665	0.041	0.221
Observations	12554	12554	12554
Adjusted R^2	0.216	0.070	0.105
Investor MSA FE	Y	Y	Y

Standard errors are in parentheses.
***p<0.01, **p<0.05, *p<0.1

Table 10: Startup Firm Success

The table shows results from the regressions of successful firm outcomes on similarity and funding type. Regression are cross-sectional at the firm level. *Initial Similarity* is the similarity score at each firm's founding year. $\mathbb{1}(Angel\ Funding)$ is an indicator variable for angel funding throughout the firm's life. $\mathbb{1}(VC\ Funding)$ is an indicator variable for VC funding throughout the firm's life. $Ln(Investments)$ is the log of the total number of investments by a given investor, $Ln(\#Funding)$ is the log of the total number of fundings that a firm receives, and $Ln(\$Funding)$ is the total funding amount that a firm receives. All specifications control for G and EG indices. Standard errors are clustered at the investor MSA level. For variable definitions and further details of their construction, see Internet Appendix C.

	Ln(#Funding) (1)	Ln($Funding) (2)	Exit - IPO (3)	Exit -IPO/Acq. (4)
Initial Similarity	0.006	0.037	-0.025	0.120
	(0.040)	(0.132)	(0.025)	(0.081)
$\mathbb{1}$(Angel Funding)	0.471***	0.542***	-0.001	0.037***
	(0.028)	(0.035)	(0.002)	(0.012)
$\mathbb{1}$(Angel Funding) × Initial Similarity	-0.724***	-2.539***	-0.211*	-0.297*
	(0.212)	(0.897)	(0.113)	(0.163)
$\mathbb{1}$(VC Funding)	0.852***	1.578***	-0.001	0.088***
	(0.010)	(0.070)	(0.004)	(0.012)
$\mathbb{1}$(VC Funding) × Initial Similarity	0.932***	6.313***	0.400**	0.317
	(0.114)	(1.259)	(0.172)	(0.210)
Ln(Firms)	0.013	-0.312***	-0.000	-0.105**
	(0.017)	(0.081)	(0.005)	(0.041)
Mean of the Dep. Var.	0.232	0.329	0.008	0.097
Observations	81448	81448	81448	81448
Adjusted R^2	0.685	0.579	0.010	0.052
Firm MSA FE	Y	Y	Y	Y

Standard errors are in parentheses.
***p<0.01, **p<0.05, *p<0.1

Internet Appendix to:

Who Finances Disparate Startups?

Figure IA.1: Funding Round Timing and Count by Investor Type

The figure shows the comparison of the frequency (left) and the number (right) of funding rounds by angel and VC investors within our sample period between 2007 and 2018.

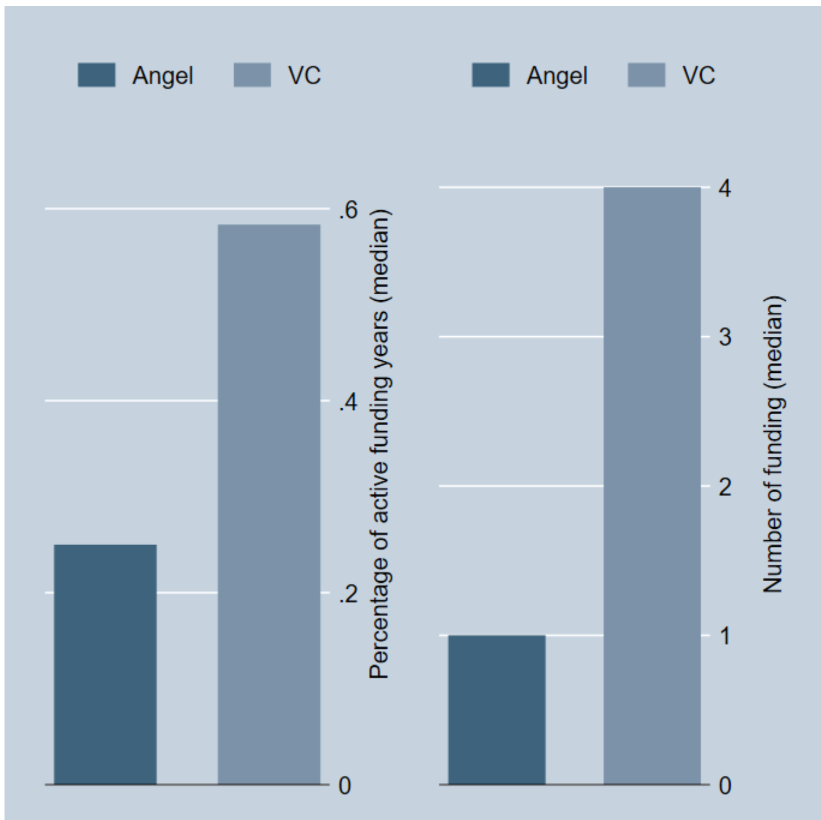


Table IA.1: Geographic Distribution of Crunchbase Firms

The table shows the geographic location distribution of firms in Crunchbase. The left panel uses the final sample firms between 2007-2018 after applying the sample selection filters described in Section 2.2. The right panel uses all firms reported with location information in Crunchbase as of April 2019.

Sample Firms				All Firms in Crunchbase			
State	Count	%	Cum%	State	Count	%	Cum%
California	33,797.00	28.3	28.3	California	62,332.00	25.6	25.6
New York	13,521.00	11.3	39.6	New York	26,890.00	11.1	36.7
Texas	7,811.00	6.5	46.1	Texas	16,038.00	6.6	43.3
Florida	6,294.00	5.3	51.4	Florida	12,406.00	5.1	48.4
Massachusetts	5,896.00	4.9	56.3	Massachusetts	11,449.00	4.7	53.1
Illinois	4,506.00	3.8	60.1	Illinois	10,065.00	4.1	57.2
Washington	3,590.00	3	63.1	Washington	7,041.00	2.9	60.1
Colorado	3,454.00	2.9	65.9	Pennsylvania	6,865.00	2.8	62.9
Georgia	3,063.00	2.6	68.5	Colorado	6,399.00	2.6	65.6
Pennsylvania	3,040.00	2.5	71	Georgia	6,329.00	2.6	68.2
Virginia	2,824.00	2.4	73.4	New Jersey	5,979.00	2.5	70.6
New Jersey	2,745.00	2.3	75.7	Virginia	5,703.00	2.3	73
North Carolina	2,237.00	1.9	77.6	Ohio	4,954.00	2	75
Ohio	2,051.00	1.7	79.3	North Carolina	4,435.00	1.8	76.8
Maryland	1,891.00	1.6	80.9	Michigan	3,953.00	1.6	78.5
Arizona	1,872.00	1.6	82.4	Maryland	3,919.00	1.6	80.1
Michigan	1,626.00	1.4	83.8	Arizona	3,886.00	1.6	81.7
Utah	1,509.00	1.3	85.1	Minnesota	3,453.00	1.4	83.1
Minnesota	1,394.00	1.2	86.2	Utah	2,993.00	1.2	84.3
Oregon	1,392.00	1.2	87.4	Tennessee	2,990.00	1.2	85.5
Tennessee	1,374.00	1.1	88.5	District of Columbia	2,888.00	1.2	86.7
Connecticut	1,208.00	1	89.5	Oregon	2,882.00	1.2	87.9
Nevada	1,182.00	1	90.5	Connecticut	2,736.00	1.1	89
District of Columbia	1,064.00	0.9	91.4	Wisconsin	2,600.00	1.1	90.1
Missouri	1,046.00	0.9	92.3	Missouri	2,556.00	1.1	91.2
Indiana	955	0.8	93.1	Indiana	2,318.00	1	92.1
Wisconsin	914	0.8	93.9	Nevada	2,217.00	0.9	93
Delaware	699	0.6	94.4	South Carolina	1,541.00	0.6	93.7
South Carolina	636	0.5	95	Kentucky	1,270.00	0.5	94.2
Kentucky	519	0.4	95.4	Alabama	1,168.00	0.5	94.7
Kansas	509	0.4	95.8	Louisiana	1,166.00	0.5	95.1
Louisiana	448	0.4	96.2	Delaware	1,140.00	0.5	95.6
Alabama	447	0.4	96.6	Kansas	1,132.00	0.5	96.1
New Hampshire	439	0.4	96.9	Oklahoma	1,099.00	0.5	96.5
Oklahoma	401	0.3	97.3	New Hampshire	969	0.4	96.9
Iowa	359	0.3	97.6	Iowa	896	0.4	97.3
Arkansas	353	0.3	97.9	Nebraska	804	0.3	97.6
Nebraska	341	0.3	98.2	Arkansas	768	0.3	97.9
Idaho	312	0.3	98.4	Idaho	630	0.3	98.2
New Mexico	269	0.2	98.6	New Mexico	617	0.3	98.5
Rhode Island	263	0.2	98.9	Rhode Island	582	0.2	98.7
Maine	237	0.2	99.1	Maine	559	0.2	98.9
Hawaii	206	0.2	99.2	Hawaii	466	0.2	99.1
Vermont	181	0.2	99.4	Vermont	381	0.2	99.3
Montana	162	0.1	99.5	Montana	341	0.1	99.4
Wyoming	149	0.1	99.6	Mississippi	328	0.1	99.5
Mississippi	113	0.1	99.7	Wyoming	258	0.1	99.7
North Dakota	89	0.1	99.8	North Dakota	248	0.1	99.8
South Dakota	85	0.1	99.9	South Dakota	233	0.1	99.8
West Virginia	78	0.1	100	West Virginia	193	0.1	99.9
Alaska	54	0	100	Alaska	172	0.1	100
Total	119,605			Total	243,237		

Table IA.2: Crunchbase Category Trends

The table presents trends of Crunchbase industry categories over our sample period between 2007 and 2018. The rank is determined by computing the percentage of firms reporting a given category each year. There are 742 Crunchbase categories, and firms can report multiple categories. The trends in category choices by firms are dominantly driven by new firms added into the data (also see dynamic similarity measure and related discussion in Section 5.4).

	Categories		
Rank	2007	2012	2018
1	Software	Software	Information Technology
2	Information Technology	Mobile	Software
3	Health Care	E-Commerce	Internet
4	Advertising	Information Technology	Health Care
5	Consulting	Health Care	Artificial Intelligence
6	Internet	Internet	E-Commerce
7	E-Commerce	Advertising	SaaS
8	Biotechnology	Social Media	Blockchain
9	Mobile	Enterprise Software	Financial Services
10	Enterprise Software	Education	Consulting
11	Manufacturing	Consulting	FinTech
12	Medical	Analytics	Machine Learning
13	Social Media	SaaS	Advertising
14	Education	Apps	Real Estate
15	Video	Biotechnology	Mobile Apps
16	Marketing	Medical	Education
17	Financial Services	Big Data	Marketing
18	SaaS	Marketing	Cryptocurrency
19	Analytics	Fashion	Mobile
20	Web Development	Finance	Apps

Table IA.3: Angel Investors

The table presents summary statistics on angel investor characteristics. There are total 7,125 unique angel investors who funded our sample firms. Panel A describes angel investors' funding amount and demographics. Panel B describes angel investors' experience in entrepreneurship. Panel C describes angel investors' reported jobs and advising positions.

	Mean	Sd	Min	Median	Max	N
Panel A: Funding and demographics						
Funding amount ('000s)	1,383.61	8,984.19	1.00	333.33	499,505.13	6,608
Seed funding amount ('000s)	343.38	520.50	1.00	206.02	10,000.00	4,206
Number of investments	3.69	8.60	1.00	1.00	243.00	7,125
Gender	0.92	0.27	0.00	1.00	1.00	7,094
Number of academic degrees	1.51	0.70	1.00	1.00	6.00	2,836
MBA (indicator)	0.24	0.43	0.00	0.00	1.00	2,836
Ph.D (indicator)	0.05	0.23	0.00	0.00	1.00	2,836
Panel B: Entrepreneurship						
Number of founded entities	1.67	0.98	1.00	1.00	8.00	3,279
% of founded entities that went public	4.88	19.17	0.00	0.00	100.00	3,279
% of founded entities that are company	84.74	32.91	0.00	100.00	100.00	3,279
% of founded entities that are investor	15.20	32.87	0.00	0.00	100.00	3,279
Panel C: Jobs						
Number of jobs (all)	6.28	7.26	1.00	4.00	67.00	6,631
Number of jobs (current)	3.63	4.48	0.00	2.00	56.00	6,631
% of jobs working as an employee	65.94	34.90	0.00	75.00	100.00	6,631
% of jobs working as a board member	26.19	27.78	0.00	20.00	100.00	6,631
% of jobs working as an advisor	2.73	9.31	0.00	0.00	85.25	6,631
% of jobs working as an executive	5.14	13.43	0.00	0.00	100.00	6,631
% of employers that are companies	69.26	40.15	0.00	100.00	100.00	6,631
% of employers that are investors	12.88	25.18	0.00	0.00	100.00	6,631
% of employers that are public companies	14.43	24.15	0.00	0.00	100.00	6,631
Number of advising roles	1.99	1.35	1.00	1.00	9.00	2,651
% of advising entities that are companies	88.56	26.72	0.00	100.00	100.00	2,651
% of advising entities that are investors	10.47	25.93	0.00	0.00	100.00	2,651
% of advising entities that are public companies	7.10	22.11	0.00	0.00	100.00	2,651

Table IA.4: Startup Similarity and Angel Funding – Funding Sample Only

The table shows results from the firm-year level regressions of funding characteristics on dissimilarity using the funding subsample. The main variable of interest is *Dissimilarity*. $\mathbb{1}(Angel)$ is a dummy variable that is equal to one if funding is received from angel investors and zero otherwise. $\$Funding$ is the total amount of funding in millions from angel investors in the given year. All specifications control for G and EG indices. Standard errors are clustered at the MSA level. For variable definitions and further details of their construction, see Internet Appendix C.

	MSA		Vintage		Soft-cosine		Dynamic	
	𝟙(Angel)	$Funding	𝟙(Angel)	$Funding	𝟙(Angel)	$Funding	𝟙(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)		
Dissimilarity	7.119***	2.015*	7.066***	2.267*	3.516***	0.419	0.427	-0.608
	(1.179)	(1.114)	(1.103)	(1.155)	(0.559)	(0.481)	(0.993)	(0.915)
Ln(Age)	-0.076***	-0.161***	-0.076***	-0.161***	-0.072***	-0.142***	-0.085***	-0.164***
	(0.020)	(0.024)	(0.020)	(0.023)	(0.026)	(0.028)	(0.021)	(0.024)
Ln(Firms)	-0.641***	-0.432***	-0.643***	-0.431***	-0.868***	-0.584***	-0.674***	-0.443***
	(0.116)	(0.080)	(0.116)	(0.081)	(0.177)	(0.123)	(0.105)	(0.071)
Observations	21178	21620	21160	21602	16590	16962	21176	21618
Adjusted R^2	0.372	0.155	0.372	0.155	0.380	0.140	0.371	0.155
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y

Table IA.5: Startup Similarity and VC Funding

The table shows the results from the firm-year level regressions of VC funding characteristics on dissimilarity. The main variable of interest is *Dissimilarity*. In Panel A and B, columns (1) and (2) use funding observations, and columns (3) and (4) use all sample observations by replacing no-funding observations with zeros. Panel B drops the entry-year observation to mitigate selection concerns. Panel C sample is restricted to first funding observations only. Columns (1) and (2) include all first funding observations, whereas columns (3) and (4) use a subsample that excludes firms that have the first funding funded by multiple types of investors. We cannot include investor-level fixed effects when the first funding is funded by more than one type of investors. $\mathbb{1}(VC)$ is a dummy variable that is equal to one if funding is received from VCs and zero otherwise. $\$Funding$ is the total amount of funding in millions from VCs in the given year. All specifications control for G and EG indices. Standard errors are clustered at the MSA level in Panel A and B and funded firm MSA level in Panel C. For variable definitions and further details of their construction, see Internet Appendix C.

Panel A: Full sample

	$\mathbb{1}$(VC) (1)	\$Funding (2)	$\mathbb{1}$(VC) (3)	\$Funding (4)
Dissimilarity	-2.370**	-338.731***	-0.144	-12.837**
	(1.182)	(87.230)	(0.181)	(6.174)
Ln(Age)	0.140***	3.526***	0.028***	0.402***
	(0.017)	(0.595)	(0.004)	(0.101)
Ln(Firms)	-0.190**	70.677***	-0.107***	1.196*
	(0.091)	(8.658)	(0.023)	(0.637)
Observations	21,178	21,620	762,299	762,299
Adjusted R^2	0.437	0.025	0.218	0.149
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y

Panel B: Dropping entry-year observations

	$\mathbb{1}$(VC) (1)	\$Funding (2)	$\mathbb{1}$(VC) (3)	\$Funding (4)
Dissimilarity	-1.986	-340.181***	0.240	-11.417*
	(1.316)	(86.732)	(0.248)	(5.903)
Ln(Age)	0.128***	9.427***	-0.002	0.654***
	(0.027)	(1.474)	(0.004)	(0.197)
Ln(Firms)	-0.170*	63.833***	-0.139***	0.746
	(0.093)	(6.587)	(0.033)	(0.514)
Observations	18155	18544	678838	678838
Adjusted R^2	0.432	0.052	0.237	0.176
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y

Panel C: First funding observations

	$\mathbb{1}$(VC)			
	All first funding		Subsample	
	(1)	(2)	(3)	(4)
Dissimilarity	0.962***	0.850***	1.075***	0.766***
	(0.139)	(0.129)	(0.171)	(0.143)
Ln(Firms)	0.044***	-0.044	0.038***	-0.078
	(0.008)	(0.067)	(0.008)	(0.070)
Observations	20,941	20,879	15,419	15,269
Adjusted R^2	0.028	0.063	0.037	0.153
Funding year FE	Y	Y	Y	Y
Founding year FEY	Y	Y	Y	
Firm MSA FE	N	Y	N	Y
Investor MSA FE	N	N	N	Y

Standard errors are in parentheses.
***p<0.01, **p<0.05, *p<0.1

Table IA.6: Startup Similarity and Entrepreneurship Quality

The table presents MSA-year-level regression of Entrepreneurship Quality Index on MSA similarity. The dependent variable is the MSA-level Entrepreneurship Quality Index measure from Andrews et al. (2019), which is available until 2013. Standard errors are clustered at the MSA level. For variable definitions and further details of their construction, see Internet Appendix C.

	EQI×100		
	(1)	(2)	(3)
MSA Dissimilarity	0.054**	0.059**	0.083**
	(0.024)	(0.026)	(0.041)
Ln(Firms)		-0.008	-0.064*
		(0.014)	(0.038)
Ln(Investors)		0.002	-0.001
		(0.002)	(0.003)
Ln(Personal Income)		0.034***	0.003
		(0.013)	(0.011)
Regional Entrepreneurship Cohort Potential Index			0.004***
			(0.000)
Regional Ecosystem Acceleration Index			-0.000***
			(0.000)
Observations	3536	3372	2054
Adjusted R^2	0.838	0.838	0.778
MSA FE	Y	Y	Y
Year FE	Y	Y	Y

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table IA.7: Startup Similarity and Angel Funding – Robustness: Form-D Filings

The table shows results from the firm-year level regressions of funding characteristics on dissimilarity. We supplement additional angel investments from the SEC Form-D filings. The main variable of interest is *Dissimilarity*. Columns (1)-(2) only use observations with any reported funding, and columns (3)-(4) use all sample observations by replacing no-funding observations with zeros. $\mathbb{1}(Angel\ Invt)$ is a dummy variable that is equal to one if angel funding is reported in either Crunchbase or Form-D filing and zero otherwise. $\$Funding$ is the total amount of angel funding (sum of Crunchbase and Form-D amounts) in millions. All specifications control for G and EG indices. Standard errors are clustered at the MSA level. For variable definitions and further details of their construction, see Internet Appendix C.

	All sample		Sub-sample	
	$\mathbb{1}$(Angel Invt)	\$Funding	$\mathbb{1}$(Angel Invt)	\$Funding
	(1)	(2)	(3)	(4)
Dissimilarity	5.060***	2.015*	0.356***	0.149***
	(1.400)	(1.114)	(0.105)	(0.054)
Ln(Age)	-0.047***	-0.161***	0.013***	-0.001*
	(0.011)	(0.024)	(0.002)	(0.001)
Ln(Firms)	-0.646***	-0.432***	-0.052***	-0.026***
	(0.095)	(0.080)	(0.012)	(0.005)
Observations	31,035	21,620	762,299	762,299
Adjusted R^2	0.396	0.155	0.169	0.093
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y

Standard errors in parentheses
***p<0.01, **p<0.05, *p<0.1

Table IA.8: Startup Similarity and Angel Funding – Robustness: Alternative Spatial Units

The table shows results reestimating the regressions in Table 6 (in Panel A) and Table 7 (in Panel B) using the alternative dissimilarity measures computed at the CSA and county levels. There are total 149 CSAs and 1,632 counties in our sample. The main variable of interest is *Dissimilarity* computed using CSA or county as the alternative spatial units. $\mathbb{1}(Angel)$ is a dummy variable that is equal to one if funding is received from the angel investors and zero otherwise. $\$Funding$ is the total amount of funding in millions from the given investor type in a year. All specifications control for G and EG indices. Standard errors are clustered at the CSA or county level. For variable definitions and further details of their construction, see Internet Appendix C.

Panel A: Dissimilarity and angel funding

	CSA		County	
	𝟙(Angel)	\$Funding	𝟙(Angel)	\$Funding
	(1)	(2)	(3)	(4)
Dissimilarity	0.599**	0.241**	0.106**	0.040**
	(0.281)	(0.116)	(0.043)	(0.019)
Ln(Age)	0.002*	-0.002	0.001	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)
Ln(Firms)	-0.027***	-0.011***	-0.015**	-0.007**
	(0.010)	(0.004)	(0.006)	(0.003)
Observations	608164	608164	748800	748795
Adjusted R^2	0.123	0.103	0.122	0.093
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y

Panel B: Robustness

	CSA			County		
	Entry-year		First funding	Entry-year		First funding
	𝟙(Angel)	\$Funding	𝟙(Angel)	𝟙(Angel)	\$Funding	𝟙(Angel)
	(1)	(2)	(3)	(4)	(5)	(6)
Dissimilarity	0.627**	0.263**	0.399***	0.140**	0.063**	0.377***
	(0.294)	(0.115)	(0.084)	(0.056)	(0.029)	(0.085)
Ln(Firms)	-0.072***	-0.029***	-0.041	-0.063***	-0.026***	-0.038
	(0.014)	(0.006)	(0.042)	(0.020)	(0.009)	(0.040)
Ln(Age)	-0.024***	-0.017***		-0.023***	-0.016***	
	(0.009)	(0.006)		(0.006)	(0.004)	
Observations	541485	541485	12907	666769	666769	15195
Adjusted R^2	0.142	0.099	0.034	0.141	0.093	0.041
Firm FE	Y	Y	N	Y	Y	N
Year FE	Y	Y	N	Y	Y	N
Funding year FE	N	N	Y	N	N	Y
Founding year FE	N	N	Y	N	N	Y
Firm MSA FE	N	N	Y	N	N	Y
Investor MSA FE	N	N	Y	N	N	Y

Internet Appendix A Theoretical Framework

The appendix shows an underlying theoretical framework on which our proposed mechanism for investor portfolio preference is based. The model shows how an investor with a geographic restriction creates her investment portfolio with startup firms. The physical position of the investor is denoted as X and the business similarity of her portfolio is S. We assume that the investor is risk neutral and that there is no strategic interaction among investors (Fulghieri and Sevilir, 2009), moral hazard (Chemmanur and Chen, 2014), and switching investor types between VCs or angels (Hellmann and Thiele, 2015).

The objective function for the investor includes the term that captures her cost of efforts to overcome the physical distance between herself and an invested firm and the knowledge gap between the industry of her expertise and the industry of the invested firm. Using a quadratic function to make the cost convex, the cost term for each project (firm) i is

$$-\{\alpha_f\left(X_i - X_f\right)^2 + \beta_f\left(S_i - S_f\right)^2\},$$

where (X_f, S_f) are the geographic position and the industry expertise of the investor, and (X_i, S_i) are the geographic position and the industry of firm i for funding. The investor cares about the relative importance between geographic distance α_f and business similarity β_f. Instead of considering a maximization of expected net payoff, we simplify the setup by directly assuming a binding constraint that the investor must spend a fixed budget on her portfolio with as many projects around (X_f, S_f) as possible.

Suppose projects are uniformly distributed in both X and S dimensions so that in each $dS \cdot dX$ area there is an equal number of projects. Then, it can be normalized to one per unit of the area. If the projects are identical except for the distance and expertise dimensions, the search-area boundary of the investor with some positive ς of the maximum cost can be written as

$$\alpha_f\left(X_i - X_f\right)^2 + \beta_f\left(S_i - S_f\right)^2 < \varsigma,$$

with the ellipse around (X_f, S_f),

$$\frac{\left(X_i - X_f\right)^2}{\varsigma/\alpha_f} + \frac{\left(S_i - S_f\right)^2}{\varsigma/\beta_f} = 1,$$

meaning that the investor searches for close and familiar projects near her location. The diameters of the ellipse are $A = \sqrt{\varsigma/\alpha_f}$ and $B = \sqrt{\varsigma/\beta_f}$, and the area inside the ellipse is

$$\pi AB = \pi\frac{\varsigma}{\sqrt{\beta_f\alpha_f}}.$$

Then, the number of projects to be financed is the same as the number of projects inside the search area, which is proportional to $(\beta_f\alpha_f)^{-1/2}$. The number of financed projects by the investor is greater when the preferences for the distance and similarity dimensions are the same as $\beta_f = \alpha_f$ (a circle). If the number of projects in any investor's portfolio is fixed as C, we can therefore derive the following condition:

$$\frac{\varsigma}{\beta_f \alpha_f} = C$$

When α_f becomes large, the geographic distance becomes more important than business expertise for angel investors relative to VCs. It follows from the above condition that β_f has to become proportionally smaller to keep $\beta_f \alpha_f$ constant. In other words, the investor with large α_f must learn about dissimilar projects when the physical distance cannot be decreased. Visually, if the search area ellipse is flatter in one dimension, it must become longer in another dimension to keep the search area the same.

The investor's tolerance for the X or S dimension along the search boundary can be found by implicit differentiation. That is,

$$\frac{d(\delta X)}{d(\delta S)} = -\frac{2\beta_f}{2\alpha_f},$$

is her tolerance for δX on the boundary when the constraint in the S dimension is relaxed by δS. We note that this tolerance measure is negative indicating that when the business dissimilarity of financed projects increases, the search must be conducted closer in the geographic distance and vice versa. The magnitude of this offsetting effect is proportional to the investor's relative sensitivity $\frac{\beta_f}{\alpha_f}$.

Next, we relax the assumption on the risk-neutral investor (*i.e.,* the investor is risk averse). It is reasonable to assume that project payoffs are correlated according to their business similarity. Then, the covariance between projects i and j becomes:

$$cov(\tilde{\theta}_i, \tilde{\theta}_j) = \mathbb{E}(\tilde{\theta}_i \tilde{\theta}_j) - \mathbb{E}(\tilde{\theta}_i)\mathbb{E}(\tilde{\theta}_j) \propto \frac{k}{|S_i - S_j|}$$

with the assumption of $\mathbb{E}(\tilde{\theta}_i) = \mathbb{E}\theta =$ constant for all i and $k >= 0$. The covariance also can be negative if we allow the negative correlation between the projects indicating a hedging effect. The variance of the portfolio of N projects with each project's cost of c_i is

$$\begin{aligned} var\left[\frac{1}{N}\sum_i(\tilde{\theta}_i - c_i)\right] &= \frac{var(\tilde{\theta}_i)}{N} + \frac{1}{N^2}\sum_{i,j} cov(\tilde{\theta}_i, \tilde{\theta}_j) \\ &= \frac{var(\tilde{\theta}_i)}{N} + \frac{(N-1)!}{N^2}\left[\frac{k}{|S_i - S_j|} - [\mathbb{E}\theta]^2\right] \end{aligned}$$

From the above, we note that the portfolio variance decreases with the dissimilarity between projects ($|S_i - S_j|$) and also with the total number of projects in the portfolio (N), while it increases with the individual project variance ($var(\tilde{\theta}_i)$). Therefore, risk-averse angel investors should take many projects and preferably different projects to diversify given the higher tolerance for S dimension.[1]

[1] It is also possible to allow that angel investors are more risk averse than VCs. If so, angel investors' preference for dissimilar projects becomes even stronger.

Internet Appendix B Crunchbase Industry Classification

Crunchbase maintains its company data using more than 700 Industries and 47 Industry Groups. See https://support.crunchbase.com/hc/en-us/articles/360043146954. Industry Groups are broader subjects that encompass multiple industries. Industries are more specific market segments. Company profiles can belong to multiple industries and industry groups.

Industry Group	Industries
Administrative Services	Archiving Service, Call Center, Collection Agency, College Recruiting, Courier Service, Debt Collections, Delivery, Document Preparation, Employee Benefits, Extermination Service, Facilities Support Services, Housekeeping Service, Human Resources, Knowledge Management, Office Administration, Packaging Services, Physical Security, Project Management, Staffing Agency, Trade Shows, Virtual Workforce
Advertising	Ad Exchange, Ad Network, Ad Retargeting, Ad Server, Ad Targeting, Advertising, Advertising Platforms, Affiliate Marketing, Local Advertising, Mobile Advertising, Outdoor Advertising, SEM, Social Media Advertising, Video Advertising
Agriculture and Farming	Agriculture, AgTech, Animal Feed, Aquaculture, Equestrian, Farming, Forestry, Horticulture, Hydroponics, Livestock
Apps	App Discovery, Apps, Consumer Applications, Enterprise Applications, Mobile Apps, Reading Apps, Web Apps
Artificial Intelligence	Artificial Intelligence, Intelligent Systems, Machine Learning, Natural Language Processing, Predictive Analytics
Biotechnology	Bioinformatics, Biometrics, Biopharma, Biotechnology, Genetics, Life Science, Neuroscience, Quantified Self
Clothing and Apparel	Fashion, Laundry and Dry-cleaning, Lingerie, Shoes
Commerce and Shopping	Auctions, Classifieds, Collectibles, Consumer Reviews, Coupons, E-Commerce, E-Commerce Platforms, Flash Sale, Gift, Gift Card, Gift Exchange, Gift Registry, Group Buying, Local Shopping, Made to Order, Marketplace, Online Auctions, Personalization, Point of Sale, Price Comparison, Rental, Retail, Retail Technology, Shopping, Shopping Mall, Social Shopping, Sporting Goods, Vending and Concessions, Virtual Goods, Wholesale
Community and Lifestyle	Adult, Baby, Cannabis, Children, Communities, Dating, Elderly, Family, Funerals, Humanitarian, Leisure, LGBT, Lifestyle, Men's, Online Forums, Parenting, Pet, Private Social Networking, Professional Networking, Q&A, Religion, Retirement, Sex Industry, Sex Tech, Social, Social Entrepreneurship, Teenagers, Virtual World, Wedding, Women's, Young Adults
Consumer Electronics	Computer, Consumer Electronics, Drones, Electronics, Google Glass, Mobile Devices, Nintendo, Playstation, Roku, Smart Home, Wearables, Windows Phone, Xbox
Consumer Goods	Beauty, Comics, Consumer Goods, Cosmetics, DIY, Drones, Eyewear, Fast-Moving Consumer Goods, Flowers, Furniture, Green Consumer Goods, Handmade, Jewelry, Lingerie, Shoes, Tobacco, Toys
Content and Publishing	Blogging Platforms, Content Delivery Network, Content Discovery, Content Syndication, Creative Agency, DRM, EBooks, Journalism, News, Photo Editing, Photo Sharing, Photography, Printing, Publishing, Social Bookmarking, Video Editing, Video Streaming
Data and Analytics	A/B Testing, Analytics, Application Performance Management, Artificial Intelligence, Big Data, Bioinformatics, Biometrics, Business Intelligence, Consumer Research, Data Integration, Data Mining, Data Visualization, Database, Facial Recognition, Geospatial, Image Recognition, Intelligent Systems, Location Based Services, Machine Learning, Market Research, Natural Language Processing, Predictive Analytics, Product Research, Quantified Self, Speech Recognition, Test and Measurement, Text Analytics, Usability Testing
Design	CAD, Consumer Research, Data Visualization, Fashion, Graphic Design, Human Computer Interaction, Industrial Design, Interior Design, Market Research, Mechanical Design, Product Design, Product Research, Usability Testing, UX Design, Web Design
Education	Alumni, Charter Schools, College Recruiting, Continuing Education, Corporate Training, E-Learning, EdTech, Education, Edutainment, Higher Education, Language Learning, MOOC, Music Education, Personal Development, Primary Education, Secondary Education, Skill Assessment, STEM Education, Textbook, Training, Tutoring, Vocational Education
Energy	Battery, Biofuel, Biomass Energy, Clean Energy, Electrical Distribution, Energy, Energy Efficiency, Energy Management, Energy Storage, Fossil Fuels, Fuel, Fuel Cell, Oil and Gas, Power Grid, Renewable Energy, Solar, Wind Energy
Events	Concerts, Event Management, Event Promotion, Events, Nightclubs, Nightlife, Reservations, Ticketing, Wedding

Financial Services	Accounting, Angel Investment, Asset Management, Auto Insurance, Banking, Bitcoin, Commercial Insurance, Commercial Lending, Consumer Lending, Credit, Credit Bureau, Credit Cards, Crowdfunding, Cryptocurrency, Debit Cards, Debt Collections, Finance, Financial Exchanges, Financial Services, FinTech, Fraud Detection, Funding Platform, Gift Card, Health Insurance, Hedge Funds, Impact Investing, Incubators, Insurance, InsurTech, Leasing, Lending, Life Insurance, Micro Lending, Mobile Payments, Payments, Personal Finance, Prediction Markets, Property Insurance, Real Estate Investment, Stock Exchanges, Trading Platform, Transaction Processing, Venture Capital, Virtual Currency, Wealth Management
Food and Beverage	Bakery, Brewing, Cannabis, Catering, Coffee, Confectionery, Cooking, Craft Beer, Dietary Supplements, Distillery, Farmers Market, Food and Beverage, Food Delivery, Food Processing, Food Trucks, Fruit, Grocery, Nutrition, Organic Food, Recipes, Restaurants, Seafood, Snack Food, Tea, Tobacco, Wine And Spirits, Winery
Gaming	Casual Games, Console Games, Contests, Fantasy Sports, Gambling, Gamification, Gaming, MMO Games, Online Games, PC Games, Serious Games, Video Games
Government and Military	CivicTech, Government, GovTech, Law Enforcement, Military, National Security, Politics, Public Safety, Social Assistance
Hardware	3D Technology, Application Specific Integrated Circuit (ASIC), Augmented Reality, Cloud Infrastructure, Communication Hardware, Communications Infrastructure, Computer, Computer Vision, Consumer Electronics, Data Center, Data Center Automation, Data Storage, Drone Management, Drones, DSP, Electronic Design Automation (EDA), Electronics, Embedded Systems, Field-Programmable Gate Array (FPGA), Flash Storage, Google Glass, GPS, GPU, Hardware, Industrial Design, Laser, Lighting, Mechanical Design, Mobile Devices, Network Hardware, NFC, Nintendo, Optical Communication, Playstation, Private Cloud, Retail Technology, RFID, RISC, Robotics, Roku, Satellite Communication, Semiconductor, Sensor, Sex Tech, Telecommunications, Video Conferencing, Virtual Reality, Virtualization, Wearables, Windows Phone, Wireless, Xbox
Health Care	Alternative Medicine, Assisted Living, Assistive Technology, Biopharma, Cannabis, Child Care, Clinical Trials, Cosmetic Surgery, Dental, Diabetes, Dietary Supplements, Elder Care, Electronic Health Record (EHR), Emergency Medicine, Employee Benefits, Fertility, First Aid, Funerals, Genetics, Health Care, Health Diagnostics, Home Health Care, Hospital, Medical, Medical Device, mHealth, Nursing and Residential Care, Nutraceutical, Nutrition, Outpatient Care, Personal Health, Pharmaceutical, Psychology, Rehabilitation, Therapeutics, Veterinary, Wellness
Information Technology	Business Information Systems, CivicTech, Cloud Data Services, Cloud Management, Cloud Security, CMS, Contact Management, CRM, Cyber Security, Data Center, Data Center Automation, Data Integration, Data Mining, Data Visualization, Document Management, E-Signature, Email, GovTech, Identity Management, Information and Communications Technology (ICT), Information Services, Information Technology, Intrusion Detection, IT Infrastructure, IT Management, Management Information Systems, Messaging, Military, Network Security, Penetration Testing, Private Cloud, Reputation, Sales Automation, Scheduling, Social CRM, Spam Filtering, Technical Support, Unified Communications, Video Chat, Video Conferencing, Virtualization, VoIP
Internet Services	Cloud Computing, Cloud Data Services, Cloud Infrastructure, Cloud Management, Cloud Storage, Darknet, Domain Registrar, E-Commerce Platforms, Ediscovery, Email, Internet, Internet of Things, ISP, Location Based Services, Messaging, Music Streaming, Online Forums, Online Portals, Private Cloud, Product Search, Search Engine, SEM, Semantic Search, Semantic Web, SEO, SMS, Social Media, Social Media Management, Social Network, Unified Communications, Vertical Search, Video Chat, Video Conferencing, Visual Search, VoIP, Web Browsers, Web Hosting
Lending and Investments	Angel Investment, Banking, Commercial Lending, Consumer Lending, Credit, Credit Cards, Financial Exchanges, Funding Platform, Hedge Funds, Impact Investing, Incubators, Micro Lending, Stock Exchanges, Trading Platform, Venture Capital
Manufacturing	3D Printing, Advanced Materials, Foundries, Industrial, Industrial Automation, Industrial Engineering, Industrial Manufacturing, Machinery Manufacturing, Manufacturing, Paper Manufacturing, Plastics and Rubber Manufacturing, Textiles, Wood Processing
Media and Entertainment	Advice, Animation, Art, Audio, Audiobooks, Blogging Platforms, Broadcasting, Celebrity, Concerts, Content, Content Creators, Content Discovery, Content Syndication, Creative Agency, Digital Entertainment, Digital Media, DRM, EBooks, Edutainment, Event Management, Event Promotion, Events, Film, Film Distribution, Film Production, Guides, In-Flight Entertainment, Independent Music, Internet Radio, Journalism, Media and Entertainment, Motion Capture, Music, Music Education, Music Label, Music Streaming, Music Venues, Musical Instruments, News, Nightclubs, Nightlife, Performing Arts, Photo Editing, Photo Sharing, Photography, Podcast, Printing, Publishing, Reservations, Social Media, Social News, Theatre, Ticketing, TV, TV Production, Video, Video Editing, Video on Demand, Video Streaming, Virtual World
Messaging and Telecommunications	Email, Meeting Software, Messaging, SMS, Unified Communications, Video Chat, Video Conferencing, VoIP, Wired Telecommunications

Mobile	Android, Google Glass, iOS, mHealth, Mobile, Mobile Apps, Mobile Devices, Mobile Payments, Windows Phone, Wireless
Music and Audio	Audio, Audiobooks, Independent Music, Internet Radio, Music, Music Education, Music Label, Music Streaming, Musical Instruments, Podcast
Natural Resources	Biofuel, Biomass Energy, Fossil Fuels, Mineral, Mining, Mining Technology, Natural Resources, Oil and Gas, Precious Metals, Solar, Timber, Water, Wind Energy
Navigation and Mapping	Geospatial, GPS, Indoor Positioning, Location Based Services, Mapping Services, Navigation
Other	#REF!
Payments	Billing, Bitcoin, Credit Cards, Cryptocurrency, Debit Cards, Fraud Detection, Mobile Payments, Payments, Transaction Processing, Virtual Currency
Platforms	Android, Facebook, Google, Google Glass, iOS, Linux, macOS, Nintendo, Operating Systems, Playstation, Roku, Tizen, Twitter, WebOS, Windows, Windows Phone, Xbox
Privacy and Security	Cloud Security, Corrections Facilities, Cyber Security, DRM, E-Signature, Fraud Detection, Homeland Security, Identity Management, Intrusion Detection, Law Enforcement, Network Security, Penetration Testing, Physical Security, Privacy, Security
Professional Services	Accounting, Business Development, Career Planning, Compliance, Consulting, Customer Service, Employment, Environmental Consulting, Field Support, Freelance, Intellectual Property, Innovation Management, Legal, Legal Tech, Management Consulting, Outsourcing, Professional Networking, Quality Assurance, Recruiting, Risk Management, Social Recruiting, Translation Service
Real Estate	Architecture, Building Maintenance, Building Material, Commercial Real Estate, Construction, Coworking, Facility Management, Fast-Moving Consumer Goods, Green Building, Home and Garden, Home Decor, Home Improvement, Home Renovation, Home Services, Interior Design, Janitorial Service, Landscaping, Property Development, Property Management, Real Estate, Real Estate Investment, Rental Property, Residential, Self-Storage, Smart Building, Smart Cities, Smart Home, Timeshare, Vacation Rental
Sales and Marketing	Advertising, Affiliate Marketing, App Discovery, App Marketing, Brand Marketing, Cause Marketing, Content Marketing, CRM, Digital Marketing, Digital Signage, Direct Marketing, Direct Sales, Email Marketing, Lead Generation, Lead Management, Local, Local Advertising, Local Business, Loyalty Programs, Marketing, Marketing Automation, Mobile Advertising, Multi-level Marketing, Outdoor Advertising, Personal Branding, Public Relations, Sales, Sales Automation, SEM, SEO, Social CRM, Social Media Advertising, Social Media Management, Social Media Marketing, Sponsorship, Video Advertising
Science and Engineering	Advanced Materials, Aerospace, Artificial Intelligence, Bioinformatics, Biometrics, Biopharma, Biotechnology, Chemical, Chemical Engineering, Civil Engineering, Embedded Systems, Environmental Engineering, Human Computer Interaction, Industrial Automation, Industrial Engineering, Intelligent Systems, Laser, Life Science, Marine Technology, Mechanical Engineering, Nanotechnology, Neuroscience, Nuclear, Quantum Computing, Robotics, Semiconductor, Software Engineering, STEM Education
Software	3D Technology, Android, App Discovery, Application Performance Management, Apps, Artificial Intelligence, Augmented Reality, Billing, Bitcoin, Browser Extensions, CAD, Cloud Computing, Cloud Management, CMS, Computer Vision, Consumer Applications, Consumer Software, Contact Management, CRM, Cryptocurrency, Data Center Automation, Data Integration, Data Storage, Data Visualization, Database, Developer APIs, Developer Platform, Developer Tools, Document Management, Drone Management, E-Learning, EdTech, Electronic Design Automation (EDA), Embedded Software, Embedded Systems, Enterprise Applications, Enterprise Resource Planning (ERP), Enterprise Software, Facial Recognition, File Sharing, IaaS, Image Recognition, iOS, Linux, Machine Learning, macOS, Marketing Automation, Meeting Software, Mobile Apps, Mobile Payments, MOOC, Natural Language Processing, Open Source, Operating Systems, PaaS, Predictive Analytics, Presentation Software, Presentations, Private Cloud, Productivity Tools, QR Codes, Reading Apps, Retail Technology, Robotics, SaaS, Sales Automation, Scheduling, Sex Tech, Simulation, SNS, Social CRM, Software, Software Engineering, Speech Recognition, Task Management, Text Analytics, Transaction Processing, Video Conferencing, Virtual Assistant, Virtual Currency, Virtual Desktop, Virtual Goods, Virtual Reality, Virtual World, Virtualization, Web Apps, Web Browsers, Web Development
Sports	American Football, Baseball, Basketball, Boating, Cricket, Cycling, Diving, eSports, Fantasy Sports, Fitness, Golf, Hockey, Hunting, Outdoors, Racing, Recreation, Rugby, Sailing, Skiing, Soccer, Sporting Goods, Sports, Surfing, Swimming, Table Tennis, Tennis, Ultimate Frisbee, Volley Ball
Sustainability	Biofuel, Biomass Energy, Clean Energy, CleanTech, Energy Efficiency, Environmental Engineering, Green Building, Green Consumer Goods, GreenTech, Natural Resources, Organic, Pollution Control, Recycling, Renewable Energy, Solar, Sustainability, Waste Management, Water Purification, Wind Energy

Transportation	Air Transportation, Automotive, Autonomous Vehicles, Car Sharing, Courier Service, Delivery Service, Electric Vehicle, Ferry Service, Fleet Management, Food Delivery, Freight Service, Last Mile Transportation, Limousine Service, Logistics, Marine Transportation, Parking, Ports and Harbors, Procurement, Public Transportation, Railroad, Recreational Vehicles, Ride Sharing, Same Day Delivery, Shipping, Shipping Broker, Space Travel, Supply Chain Management, Taxi Service, Transportation, Warehousing, Water Transportation
Travel and Tourism	Adventure Travel, Amusement Park and Arcade, Business Travel, Casino, Hospitality, Hotel, Museums and Historical Sites, Parks, Resorts, Timeshare, Tour Operator, Tourism, Travel, Travel Accommodations, Travel Agency, Vacation Rental
Video	Animation, Broadcasting, Film, Film Distribution, Film Production, Motion Capture, TV, TV Production, Video, Video Editing, Video on Demand, Video Streaming

Internet Appendix C Variable Definition

Variable Name	Definition
Similarity	A firm-MSA-year level business similarity measure computed by taking the average of cosine similarity scores between the focal firm and each of the rest of the firms in a given MSA-year using 742 Crunchbase industry categories.
Industry Similarity	A within-investor industry similarity of portfolio firms.
Angel	An indicator variable equal to one if given investor's Crunchbase investor type includes angel and zero otherwise.
VC	An indicator variable equal to one if given investor's Crunchbase investor type includes venture capital and zero otherwise.
Ln(Firms)	The natural logarithm of the number of firms in Crunchbase.
Ln(Investors)	The natural logarithm of the number of investors in Crunchbase.
Ln(Net Jobs)	The natural logarithm of the number of net jobs created in an MSA from the Business Dynamics Statistics (BDS) from the public Census data.
LN(Firm Deaths)	The natural logarithm of the number of firms that have exited in their entirety.
Ln(Personal Income)	The natural logarithm of per capita personal income obtained from the Bureau of Economic Analysis.
Entrepreneurship Quality Index (EQI)	MSA-year level average entrepreneurial quality obtained from Andrews et al. (2019). The data is available between 1988-2016.
Regional Entrepreneurship Cohort Potential Index (RECPI)	MSA-year level expected number of growth events given the start-up characteristics of a cohort at birth obtained from Andrews et al. (2019). The data is available between 1988-2016.
Regional Ecosystem Acceleration Index (REAI)	MSA-year level ratio of the realized growth events to expected growth events (RECPI) obtained from Andrews et al. (2019). The data is available between 1988-2013.
Treated	Treated is a continuous variable and measures the fraction of families that may have lost accreditation due to the act at the state level using the enactment of Dodd-Frank Act as an exogenous shock to the number of accredited investors.
Post	An indicator variable, which is equal to 1 if after the enactment of Dodd-Frank in 2010 and 0 otherwise.
High Diversity	A MSA-year level angel diversity indicator variable equal to 1 if the diversity measure (*Education* or *Demographics*) is above the median and 0 otherwise. Each diversity measure is computed as the sum of normalized diversity component by its mean and standard deviation, where each component is calculated as $1 - HHI(component)$. *Education* consists of college and major, and *Demographics* consists of gender and race (Asian-pacific, Black, and White). HHI(component) is computed as following: For the MSA with three angel investors, where two majored in computer science and one majored in literature, the HHI(major) is equal to 0.56 $((1/3)^2 + (2/3)^2)$.
Investor Type	An indicator variable equal to one if funding is raised from a given investor type and zero otherwise.
%Investor Type	The average fraction of the funding amounts received from the given investor type, weighted by the size of each funding round in a year.
%Funding	The fraction of the total funding amounts received from the given investor type in a year.
$ Total Funding	The total funding amount received in a year (in millions).
Geographic Distance	A distance in the mile between an investor and funded firm.
Ln(Investments)	The natural logarithm of the total number of investments of a given investor.
Ln(Investing Years)	The natural logarithm of the number of years since the first investment record.
Subsequent Funding	An indicator variable equal to one if a given firm has a record of any type of subsequent funding.
Exit Rate-IPO (Acq)	The average IPO (Acquisition) exit rate among all funded firms for a given investor.
Close	An indicator variable equal to one if an investor's portfolio firms consist of below median average distance between the investor and portfolio firms and zero otherwise.
Dissimilar	An indicator variable equal to one if an investor's portfolio consists of firms with below median average similarity and zero otherwise.
Close&Dissimilar	An indicator variable in the intersection of *Close* and *Dissimilar*
G and EG Index	The indices are computed following Ellison and Glaeser (1997). The indices are measures of the geographic concentration of an industry, based on a location choice model considering the localized industry-specific spillovers, natural advantages, and pure random chance of plant location choices. The G Index is computed as the sum of squares of the difference between the observed concentration of state-industry employment beyond the model estimate, and the EG index further controls for the differences in the size of the distribution of plants and the size of the geographic areas.