Who Finances Disparate Startups?

S. Katie Moon Paula Suh*

Abstract

We examine the novel association between local startup industry composition and angel financing. Using Crunchbase data, we find that angel investors contribute to local startup industry diversity by funding startups dissimilar to incumbent ones. Angel investors attain industry diversity, particularly by financing startups in emerging industries and engaging with local angel investors from diverse backgrounds. Our findings are driven by angel investors' unique portfolio preference for geographic proximity to funded firms and industry diversity while venture capitalists typically lean toward industry specialization. Angel investors' investment behavior is consistent with reinforcement learning, where exploratory investment strategies are optimal for less experienced investors.

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

JEL Classification: M13, L26

*Moon is at the 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, the University of Georgia, FOM Research Group Virtual Corporate Finance, and conference participants at the HEC-McGill Winter Finance Conference, the EFA meeting, the FMA meeting, the Colorado Finance Summit, CICF, as well as Hadiye Aslan, Tetyana Balyuk, Andrea Buffa, Maria Chaderina, Susan Chaplinsky, Michael Ewens, Laurent Fresard, Jack He, Thomas Hellmann, Shenje Hshieh, Yael Hochberg, Pekka Honkanen, Thomas Krause, Steven Malliaris, Yifei Mao, Gordon Phillips, Merih Sevilir, Shrihari Santosh, Antoinette Schoar, Ilya Strebulaev, Katrin Tinn, Sheridan Titman, Heather Tookes, Yuri Tserlukevich, Xinxin Wang, 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

Agglomeration has long been recognized as an important driver of economic growth through the concentration of an industry (Marshall, 1890). The benefits of industry agglomeration arise from input sharing, improved matching in local labor markets, and knowledge spillovers (Jaffe, Trajtenberg, and Henderson, 1993; Rosenthal and Strange, 2008; Kolympiris, Kalaitzandonakes, and Miller, 2011).[1] These benefits have persisted, explaining the success of Silicon Valley as the nation's leading startup hub for many years. However, a recent trend in startup formation, especially in high-tech industries, exhibits starkly different characteristics from those of Silicon Valley. There is now a greater geographic dispersion of startups across smaller cities in the United States, and, more importantly, these rising hubs appear to embrace more diverse industries.[2]

In this paper, we examine how angel investors contribute to the recent rise of startup hubs with greater industry diversity in smaller cities through their entrepreneurial financing. Angel investors have become an important source of entrepreneurial financing with a substantial impact on startups. First, angel investors are the primary source of early-stage equity financing, where the aggregate amount of angel funding is as big as that of VC funding (Hellmann and Thiele, 2015, 2019; Hellmann, Schure, and Vo, 2021). Second, they improve the survival, success, and growth of startups (Kerr, Lerner, and Schoar, 2014; Hellmann and Thiele, 2019). Third, angel investors have a wide geographic presence (Angel Capital Association, 2017) unlike VCs, who tend to concentrate in a few areas like Silicon Valley, New York, and Boston (Chen, Gompers, Kovner, and Lerner, 2010).

The growing role of angel investors appears to be associated with the emergence of startup hubs in smaller cities. Table 1 presents a list of Metropolitan Statistical Areas (MSAs) where

[1] Marshall (1890) emphasizes that agglomeration ultimately reflects gains that reduce the costs of moving goods, people, and ideas. For more recent examples, see 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."

angel investors have rapidly increased over the 2007–2018 period. The top 30 MSAs with the highest growth in the number of angel investors largely overlap with the recent rising startup hubs often highlighted by the media, such as Denver, Seattle, Durham, and San Diego. More importantly, these MSAs with significant growth in angel investors exhibit increasing industry diversity, as measured by the growth rate of startup business similarity (our main variable of interest, described in Section 3). Based on these observations, we aim to investigate a novel, previously unexplored link between the industry composition of local startups and the types of early-stage financing using a comprehensive dataset on a large number of startups and their granular industry and financing information.

[Insert Table 1 Here]

We use Crunchbase as our major data source. It is the largest leading platform for superb business and funding information on early-stage firms and their financiers. Importantly, while most private firms lack publicly available industry classification data, such as SIC codes, Crunchbase tags firms with its own industry classification, comprising 47 broad industry sectors and 742 detailed industries. This rich classification allows us to create a detailed industry vector of 742 elements for each firm. To assess regional startup industry diversity, which is the key focus of our paper, we compute a cosine-similarity measure of these industry vectors across all startups within an MSA. The similarity measure evolves over time as the composition of geographic peers changes with the entry of new firms and the closure of existing ones.

Our central research question examines whether angel investors contribute to forming a more diverse local group of startups. To establish a causal interpretation of the effect, we exploit the regulatory change in 2013 that permitted general solicitation of private offerings. Permitting general solicitation of private offerings means advertising an active capital raise to a broad audience, effectively encouraging participation from general individual investors (*i.e.,* angel investors) in private capital raising that would have otherwise been limited to specific, often well-connected groups of investors (*i.e.,* VCs). We find that an increase in

the participation of angel investors in private offerings, facilitated by this regulatory change, leads to an 11.6% increase in local startup industry diversity accounting for more than half of the average change in industry diversity in our entire sample. This finding supports the idea that local industry diversity is likely associated with angel financing.

We then show that startups exhibiting greater dissimilarity from their geographic peers are indeed more likely to receive angel financing. Specifically, a one standard deviation increase in a firm's *dis*similarity to its geographic peers increases the likelihood of receiving angel financing by 39.3% and the amount of angel funding by 37.5%. We do not observe similar patterns for VC financing, as an increase in business dissimilarity rather *decreases* startups' access to VC financing. These findings highlight angel investors' unique investment styles and the distinct role of their financing in influencing local startup dynamics.

We further address potential concerns about alternative explanations through a battery of robustness tests. Our results remain robust when using alternative business similarity measures, which account for a possible mechanical decrease in similarity score due to expanding industry categories (termed vintage similarity) and for correlations among sub-classifications within broad industry sectors (termed soft-cosine similarity). The year-to-year variation in our business similarity measure arises only from the simultaneous entry or exit of startups in an area over time. However, we also conduct our analysis after removing the entry year observations to further strictly isolate the exogenous changes in business similarity from its potential correlation with a firm's initial entry decision. Additionally, we alleviate the possibility of spurious correlations. First, we supplement the angel investment data with Form-D filings from the Securities and Exchange Commission (SEC) to address concerns about potential missing angel investment reports in Crunchbase. Second, we obtain the detailed time stamps of reported funding incidences in Crunchbase and remove any lagged funding reports to mitigate the concern about selection bias in reporting only successful funding. Lastly, we show that our results are not driven by the correlation between business similarity and startup quality.

Angel investors' portfolio preference for diverse startups is not a prevalent investment pattern among all entrepreneurial financiers and is especially distinct from other entrepreneurial financiers, especially venture capitalists (VCs), who comprise another prominent group of early-stage financiers. Specifically, angel investors' portfolios consist of startups that are significantly more diverse in industry but geographically closer to angel investors than those of VCs. This implies that angel investors tend to rely on physical proximity for deal sourcing while preferring to fund startups from diverse industries, given their geographic constraints.

We explain this distinct investment pattern of angel investors through a bandit model framework. In a multi-armed bandit model with reinforcement learning, an agent determines a transition from exploring to exploiting arms by facing a trade-off between increasing cumulative rewards and discovering the optimal arm as she accumulates experience. Under this framework, less-experienced angel investors are expected to explore startups from a wide range of industries, similar to agents in the exploration phase, whereas more-experienced VCs tend to focus on specific industries with higher mean awards based on their accumulated knowledge and experience. Furthermore, the model suggests that an angel investor can gradually behave like VCs over time, specializing in a small number of industries as her experience grows. We find these patterns in our data and further substantiate the idea of reinforcement learning by showing varying degrees of distinctive funding preferences even within angel investors across different investment experiences and sizes. Importantly, this reinforcement learning mechanism is not driven merely by investors' preference for startups in certain stages (*e.g.,* VCs generally invest in more mature startups than angels) but rather by their investment experiences.

Next, we turn our attention to individual angel investors to enrich the interpretations of our main results by exploring the types of startups that angel investors fund. We find that angel investors promote local industry diversity by being more willing to explore and fund startups in new and emerging industries, which may face greater capital market frictions due to relatively high uncertainties in their ultimate success. We then investigate whether the

individual demographic diversity of local angel investors affects their decisions to fund more diverse startups. Given the geographic constraint, angel investors collect information locally and participate in local angel groups to learn about investment opportunities and improve their investment decisions (Cable, 2011). In line with this reasoning, we find that local angel groups with more diverse demographic and educational backgrounds are marginally more likely to fund disparate startups. This result has important policy implications, suggesting that fostering diversity among individual investors in a region can be a driving force in nurturing unique business ideas among startups.

Finally, we discuss the broader implications of angel investors' preference for funding diverse startups on their portfolio outcomes and how it, in turn, influences startups' birth location decisions.[3] In a cross-sectional test of angel investors' portfolio outcomes, we find that angel investors with more diverse funding portfolios achieve higher success rates, as measured by subsequent funding rounds and exit rates through IPOs or acquisitions. This finding is consistent with the implication of the bandit model, where funding diverse startups accomplishes optimal exploration given angel investors' relatively lower investment sophistication or experience compared to VCs. More importantly, angel investors appear to play an important role in locally attracting more diverse startups. In a firm-level cross-sectional analysis of birth location choices, we find that a startup is 3.4% more likely to enter an MSA with greater angel investor participation if it can differentiate themselves from the local incumbent startups.[4] This suggests that startups consider the funding preference of angel investors when deciding on their birth locations, as they seek to obtain angel funding.

Collectively, our results underscore the significant role of angel investors in locally funding and fostering a more diverse group of startups. Given their widespread presence across the country, angel investors attract disparate startups by influencing their entry decisions with the early-stage funding they provide and, thus, facilitate the formation of geographi-

[3]Unlike Bryan and Guzman (2023), who examine startups' relocation after the initial founding, we focus on startups' birth locations.

[4]Importantly, we show later that our results are not driven by the possibility that local colleges foster entrepreneurship in the area.

cally dispersed startup hubs in smaller cities. We also highlight that the rising hubs embrace the diversity among both startups and individual angel investors, serving as a catalyst for regional economic growth (Jacobs, 1969). In contrast to Silicon Valley, which is often described as a prime example of regional economic growth driven by industry concentration and specialization (Marshall, 1890; Arrow, 1962; Romer, 1986), these new hubs in smaller cities represent a departure from the traditional agglomeration economics with their emphasis on industry diversity instead.

Our paper relates to three areas in the economics and finance literature. First, our paper adds to the entrepreneurial agglomeration literature. The recent agglomeration literature expands Marshall (1890) into studies of highly localized entrepreneurial clustering driven by intellectual spillover (Rosenthal and Strange, 2008; Glaeser et al., 2010; Kolympiris et al., 2011; Carlino and Kerr, 2015; Chatterji et al., 2014). Also, previous studies have extensively examined VCs as an important factor of entrepreneurial agglomeration, typically locating near tech hubs and attracting new startups in proximity (Chen et al., 2010; Hochberg, Ljungqvist, and Lu, 2007, 2010). For example, Chen and Ewens (2023) use regionally disproportionate shocks to the supply of VC funds to show spillover effects on local startups' financial constraints and location choices. Our paper makes a unique contribution to the literature by documenting that angel investments are one of the important mechanisms driving industry diversity in geographically dispersed regions and fostering diversity-driven local economic growth (Jacobs, 1969). Importantly, we emphasize that had the geographically dispersed formation of startups been driven by VC investors, we would observe an increase in industry concentration rather than industry diversity.

Second, our paper closely relates to the literature on angel investors. Despite the growing role and size of angel financing (Hellmann and Thiele, 2019; Hellmann et al., 2021), the research on angel investors is underdeveloped compared to the extensive work on VC investors likely due to data limitations. Previous studies on angel investors focus on the real effects of angel financing and the success of angel-financed startups. Kerr et al. (2014) show that

angel investments enhance startups' survival and performance but not their likelihood of future funding. Lindsey and Stein (2019) show that a reduction in a market's pool of angel investors decreases firm entry and local employment, whereas Denes, Howell, Mezzanotti, Wang, and Xu (2020) find no significant effect of angel investor tax credits on entrepreneurial activity. Our study distinctively focuses on angel funding to startups that are *dissimilar* from local incumbents but also complements prior studies by showing evidence for company-led financing choice (Hellmann and Thiele, 2019) and proposing local startup business diversity as a novel factor in angel investors' funding decisions (Bernstein, Korteweg, and Laws, 2017).

Lastly, our paper broadly connects to the literature on early-stage financiers' portfolio choices. We highlight how angel investors' preference for geographically close startups from diverse industries largely contrasts with that of VC investors, who typically favor industry specialization (Sorenson and Stuart, 2001; Hochberg et al., 2007; Chen et al., 2010). We explain these contrasting portfolio preferences using the bandit problem framework by distinguishing between angel investors and VCs based on their experience levels in learning about the success probability distribution of different options. We further relate these portfolio preferences of angel investors vs. VCs to startup firms' entry decisions.

2 Data and Sample

2.1 Crunchbase Data

We put together our detailed data on startup firms and their funding information from Crunchbase. Crunchbase is a crowdsourcing platform that 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 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 Crunchbase 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 all 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 has 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 use these identifiers to connect firms to their funding rounds and investor information. Besides, Crunchbase also provides detailed information on exits, including initial public offerings (IPOs) and acquisitions, as well as founders and team members.

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. The similarity score is a dynamic measure and is simultaneously determined by incumbents and other newly entering or closing firms in the area. This makes it impossible to target a specific similarity score; hence, a firm's choice to be on Crunchbase does not seem to correlate with our similarity measure in any clear way. Also, being on Crunchbase does not guarantee access to financing or any specific type of financing. We find that 15% 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

Using Crunchbase, we create comprehensive data on startups and their funding information. We restrict our sample to startups that have a non-missing headquarters location, valid zip code, and founding year. There are some very old public and private firms in Crunchbase. To focus solely on the early-stage startups, we use the median age (from their founding) at the IPO of 2,067 exiting firms in our sample, which is ten, as the age cutoff. After these screenings, our final sample comes down to 117,944 unique startups founded between 2007 and 2018 with over 700,000 firm-year observations.[5] We then add the investor and funding information to the final sample. Crunchbase reports various funding types, which we categorize into three broad groups: Angel, VC, and Others. Others include funding provided by investment banks, hedge funds, pension funds, private equities, and accelerators. About 15% of the firms in our sample have at least one funding record (denoted as funding sample) and corresponding funding investor information. The funding amount of no funding record

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

is replaced with zero, which means the firm has not received any funding that year.

Panel A of Table 2 describes firm age and funding information for the funding sample. The average (median) age of firms in our funding sample is 3.81 (3). 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. 3.38 years old) when they receive angel financing than VC funding. 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 funding is smaller than that of VC funding ($875,000 vs. $4.3 million).[6] Angel investors tend to fund firms in significantly closer geographic proximity compared to VC investors (88 miles vs. 427 miles). The distance is measured using zip codes, and the observations with zero miles distance mean the investor and funded firm are located in the same zip code.

3 Business Similarity Measure

One unique feature of Crunchbase is its own 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 granular industry classifications of its own. There are 46 broad industry sectors, which cover anything from Consumer Goods to Information Technology, and a total of 742 finer subcategories (see Internet Appendix A for the 742 industries and their groups). For example,

[6]These are total amounts per funding round, each of which includes on average 2 and 2.41 investors for angel and VC funding, respectively, as shown in the table.

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 choose multiple categories that describe their business at the time of creating a company profile and are allowed to update them over time.[7] In our sample, firms report, on average, three categories. Only about 10% of firms report more than 5 categories.[8]

Using the detailed industry classifications provided by Crunchbase, we create a firm-year-level business similarity measure. 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 industry 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 startup 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 classification by Crunchbase includes the category h and zero otherwise. Then, we 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

[7] In Table IA.2, we provide a list of the top 20 popular industry categories chosen by firms over time. The list ensures that the Crunchbase industry 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 recent years. We discuss the implications of the expanding categories and our adjusted similarity scores in detail later in this section.

[8] Based on our conversation with a Crunchbase representative, it is the firms who select the categories without any input from Crunchbase.

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 the existing 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 (*i.e.,* industry concentration).

Panel B in Table 2 describes our similarity score measure. The average and median similarity scores of our sample firms during the sample period are 0.0376 and 0.0309, respectively, with a standard deviation of 0.0305.[9] Although the levels of similarity scores are small, this is due to the mechanical reason that Crunchbase maintains a wide array of industry categories in our cosine similarity calculation, while firms, on average, choose only three industries out of 742 available categories. However, we note that the changes in similarity score levels, which are the focus of our paper, are economically significant. Over the sample period between 2007 and 2018, the average similarity score decreased by about 0.007 (from 0.041 to 0.034), which is equivalent to a 17.1% change.

We further illustrate magnitudes of similarity scores and their variations with a few notable examples to help with the interpretation of our results. The most intuitive comparison would be Silicon Valley (San Jose-Sunnyvale-Santa Clara MSA), whose similarity score is 0.0487. As expected, the Silicon Valley area's similarity score is above the sample average as it is comprised of similar technology-based firms. One standard deviation change of 0.0305 in similarity scores would be equivalent to comparing Silicon Valley and Modesto, California. 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

[9]The MSAs with sample average similarity scores are Tulsa, Oklahoma (0.0382), Burlington, Vermont (0.0382), and Cleveland-Elyria-Mentor, Ohio (0.0385).

from winery and food producers to bottle manufacturing and steelworking companies.[10]

Next, we lay out a few potential concerns with the nature of Crunchbase industry classifications. One concern is that our similarity score may mechanically decrease over time due to an increase in the number of firms or the expansion of the total number of Crunchbase industry categories. First, we do not find that an increase in the number of firms monotonically decreases the similarity score. Despite the number of firms steadily increasing in most MSAs over time, similarity scores move up and down, indicating that the increase in the number of firms does not simply result in a decrease in similarity scores (see Figure IA.1). Second, the concern about the expansion of Crunchbase industry categories is valid, as two expansions were made in the number of categories during our sample period.[11] Fortunately, we have time-stamped information on when each industry category is created and updated. To address the concern that the expansion of Crunchbase industry categories may mechanically decrease our similarity score over time, we create an additional measure called "vintage similarity score." This measure uses only the 419 categories that existed from the beginning of our sample period in 2007. For example, newly created categories in the middle of our sample period, such as "Blockchain" and "FinTech," are completely excluded from the analysis based on this measure. This approach ensures that any decrease in the vintage similarity score is not mechanically driven by the addition of new categories over time.

We also consider the possibility of overstating the dissimilarity due to the granularity of Crunchbase industry classifications. Some finer sub-categories are closely related within a broader industry sector than others. Therefore, we address this concern by creating an additional measure called "soft-cosine similarity score." This measure accounts for any correlations among finer sub-categories within 46 broad industry groups (see Internet Appendix A), based on the soft-cosine similarity calculation technique introduced by Sidorov, Gelbukh,

[10]Texarkana, Texas, has the highest similarity score of 0.315 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. Gulfport-Biloxi, Mississippi (0.0010), Winchester, Virginia (0.0070), and Sheboygan, Wisconsin (0.0086) are the areas with the lowest similarity scores.

[11]Some categories were also removed, although substantially fewer categories were removed than added.

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 sector to be treated as equally more similar.[12] We assume a 5% correlation between the elements under the same broad industry sector, and our results largely hold with using a higher or lower correlation. This approach ensures that we are not overly loading on the difference between any two categories that are potentially related under the same broad industry sector. We use our three different similarity scores throughout all our empirical analyses to deliver consistent results and clear interpretations.

4 Angel Investors and Identification Strategies

4.1 Characteristics of Angel Investors

Angel investors have become a growing source of early-stage financing. In Figure 1, we show a dramatic increase in both the cumulative number of angels and the amount of angel funding (Panel A) and angel investors' regional presence (Panel B) during our sample period. We also provide a detailed summary statistics of angel investors in Table IA.3. 7,125 unique angel investors have funded firms in our sample. It is notable that Crunchbase data shows a sizable heterogeneity in angel investor characteristics and their funding amounts. It captures a wide spectrum of angel investors, from individuals to well-known ex-entrepreneurs with large wealth and more experienced large angel groups, indicated by a wide range of funding amounts from as small as $1,000 to $500 million. They make, on average, 3.69 investments during our sample period. Angel investors are dominated by male investors and, of the reporting sample, 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

[12]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.

experience, they have found, on average, 1.67 companies.

[Insert Figure 1 Here]

Prior studies document stylized facts on angel investors in comparison to VCs. Wharton Entrepreneurship and Angel Capital Association (2017) reports that angel investors are geographically dispersed relative to VCs. 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 geographic proximity to their homes: 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 network of syndicates with a larger investment size allows VCs to specialize in a small number of industries (Hochberg et al., 2010; Hochberg, Mazzeo, and McDevitt, 2015; Lerner, 2020) while investing more frequently in spatially distant companies as the VC networks diffuse deal information across boundaries and expand the spatial radius of their search (Sorenson and Stuart, 2001; Hochberg et al., 2007).

The angel investor and VC characteristics described using our Crunchbase data align with the stylized facts above. In Table IA.4, we examine the business similarity among funded firms and the average geographic distance between each investor and the firms in their funding portfolio. We observe that angel investors' portfolios span startups from more diverse industries compared to portfolios of other types of investors, particularly VCs. Angel investors also exhibit a significantly stronger preference for geographic proximity to their funded firms than other investors.[13] Furthermore, when examining funding-level observations, we find that angel investors are more likely than other types of investors to add new

[13]The test in Table IA.4 is cross-sectional across investors, with each observation representing one investor. We exclude investors with only one funded firm from the analysis, as business similarity cannot be computed in such cases. Importantly, these results are not driven by differences in the total number of funded firms between angel and VC portfolios, as our results remain strong after including the fixed effects for the investor-level total number of investments. In Table IA.5, we conduct an analogous test within matched investment count bins and confirm that the findings remain robust and consistent.

firms that are in dissimilar industries but geographically close to them. These results confirm that angel investors prefer industry diversity within geographic constraints, while VCs lean toward industry specialization by relying more on their expertise and experience.

The different portfolio preferences of angel investors and VCs do not necessarily imply that they are perfect substitutes for funding early-stage startups (Hellmann et al., 2021). Our results can be interpreted in the context where successful angel-backed startups in earlier stages migrate to VC funding in later stages (Chemmanur and Chen, 2014). As such, angels typically invest in earlier-stage, smaller companies with a broader industry focus, whereas VCs tend to invest in later-stage, high-growth firms within specific sectors. Our paper particularly emphasizes angel investors' critical role in sustaining and nurturing startups that may be initially overlooked by VCs (*e.g.,* startups in emerging industries) due to their industry specialization, ultimately enabling these firms to develop into successful ventures that attract VC investment later on. We also note that these different portfolio preferences between angel and VC investors documented in this paper extend beyond their focus on startups in different stages. Instead, we contend that such different preferences for industry diversity stem from a wide spectrum of investor experience and knowledge, which cannot be explained by their different focus on investment stages. We further elaborate on this idea in the next section.

4.2 Investment Strategies of Angel Investors

In this section, we further attempt to explain the differences in portfolio characteristics between angel investors and VC investors using the multi-armed bandit problem, which delineates the exploration-exploitation trade-off in reinforcement learning. The concept of multi-armed bandit models has been extensively applied in studies of decision-making through learning. For instance, March (1991) examines organizational learning based on the tension between the exploration of new possibilities and the exploitation of old certainties. Roberts and Weitzman (1981) study optimal funding criteria between two types of projects

that require exploitation vs. exploration, which can be generalized into the general theory of information-gathering processes. Moscarini and Smith (2001) heavily rely on a bandit model to describe the optimal level of experimentation. Also, Manso (2011) embeds bandit problems into a principal-agent framework to investigate incentives for exploration and exploitation in innovation.

A multi-armed bandit model consists of K (where $K \geq 2$) arms, each associated with an unknown probability distribution P_i with a mean μ_i. In our context, the K arms represent potential startup firms across K different industries within a given physical distance that an agent can consider for investment. P_i represents the success probability of startup firm i. Pulling arm i represents that the agent invests in firm i, yielding a reward r sampled from P_i. P_i is agent-specific considering the monitoring and advising role of early-stage investors (Gompers, Gornall, Kaplan, and Strebulaev, 2020; Gornall and Strebulaev, 2021; Hochberg et al., 2007, 2010), thereby precluding a mimicking strategy. The agent has a budget of T arm pulls within her lifetime and sequences these pulls to maximize the accumulated reward after T pulls. The reward for pulling arm i at time t is r_{it}. The agent's goal is to maximize $\sum_{t=1}^{T} r_{it}$, where $r_{it} \sim P_i$, in her lifetime. The agent can pull multiple arms at a time depending on her budget for each round. For simplicity, pulling multiple arms at a time can be considered multiple-arm pulls without intervals.

The arm with the highest mean reward is termed the optimal arm. Initially, the agent is unaware of the reward distributions of each arm, including the optimal arm, which necessitates exploration by pulling the arms in sequence to learn these distributions. Subsequently, the agent exploits by pulling the arm believed to be the most rewarding (the optimal arm) to maximize cumulative reward. This constitutes the exploration-exploitation trade-off in reinforcement learning, where the agent faces a dilemma of accumulating less reward by exploring too much or missing out on discovering the optimal arm by exploiting too soon. Various solutions to the multi-armed bandit model, such as ϵ-Greedy Selection (Watkins, 1989), Upper-Confidence-Bound Arm Selection (Auer, Cesa-Bianchi, and Fischer, 2002),

and Thompson Sampling Strategy (Thompson, 1933; Agrawal and Goyal, 2012), suggest the optimal timing to transition from exploration to exploitation.

We posit that such a transition point from exploration to exploitation represents the differing investment styles of angel investors and VC investors. Before transitioning to an exploitation strategy, the investor explores by pulling as many arms as possible to identify the optimal one. Therefore, investors, especially those less experienced in early-stage entrepreneurial financing, are more likely to behave similarly to agents in the exploration phase of the described bandit model. Conversely, more experienced investors who have a better understanding of which industries offer higher mean rewards given their characteristics are likely to switch to an exploitation strategy, focusing on specific industries. We expect the former to be angel investors and the latter to be VC investors based on their contrasting stylized characteristics we described in Section 4.1. In this setting, VCs enter as more experienced investors who have already identified their optimal industries for investment as financial institutions, whereas angels enter as relatively less experienced investors as individuals but grow toward more experienced investors over time. Consequently, we predict that angel investors evolve to adopt investment styles similar to VC investors as they gain experience and expand their investment size. These predictions are examined in Figure 2.

[Insert Figure 2 Here]

We use the wide spectrum of our angel investor characteristics, ranging from a single person to a large group of angels who resemble micro-VCs, and classify them as more or less experienced based on the number of past investment experiences in Panel (a) and investment size in Panel (b). We then plot the characteristics of their portfolios with the geographic distance on one axis and industry specialization on the other, by one of the proxies for investing experience (represented by the marker size). We first find in Panel (a) that the business diversity among funded firms in angel investors' portfolios starts high but decreases as the investment sequences rise. For example, the business diversity among the funded firms in an angel investor's investment sequence beyond her fifth funding is approximately 6% lower

than that of her first two investments. In Panel (b), we also find that funded firms in angel investors' portfolios become more similar in industries as their investment size increases. The business diversity when the investment size of an angel investor is sufficiently large, for example, in the 10th decile, is approximately 7% lower than that when the investment size is in the second decile. The results in both panels imply that as angel investors gain more experience, as measured by the investment sequence or size, their portfolio preferences increasingly resemble those of VCs, characterized by industry specialization.

Overall, our bandit model framework plausibly depicts the stylized facts on the different portfolio preferences of angels vs. VCs. We propose the reinforcement learning of angel investors by exploring various industries as the mechanism behind these contrasting portfolio preferences. The findings in this section suggest that the business similarity of local startups is an important consideration for investors' funding decisions and that angel financing is likely associated with the increased business diversity in entrepreneurship in a region.

4.3 Identification Strategy

In this section, we consider a regulatory shock to angel investors' participation to show the causal effects of angel financing on local startup dynamics. We specifically consider the elimination of the prohibition of general solicitation under the Jumpstart Our Business Startups (JOBS) Act, implemented on September 23, 2013.

Startup firms' capital raising typically relies on Rule 506 of Regulation D, which provides a safe harbor based on the exemption in Section 4(2) of the Securities Act for sales of private (unregistered) securities sold to accredited investors. The sale of such private securities 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 provide funding to startups in exchange for private securities are generally accredited investors. For these private securities to be exempt from the regulatory requirements that public securities have, the issuers must not use general solicitation to market the securities (*i.e.,* the prohibition

against general solicitation and advertising in Rule 506). For example, under the prohibition of general solicitation, any issuers relying on Rule 506 of Regulation D exemptions were disallowed from advertising the sale of securities in newspapers and magazines, on public websites, or through communications broadcast over television and radio.

However, Section 201(a) of the JOBS Act (henceforth, the general solicitation shock) sought to reduce barriers to capital formation, particularly for smaller companies, and amended Rule 506 of Regulation D to permit solicitation or advertising if the sale of private securities is limited to 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 the capital raise that would have otherwise been limited to particular groups of investors (*i.e.,* VCs). This change also reduces the search frictions between entrepreneurs and angel investors. According to the Director of the Division of Corporation Finance at the SEC, over 900 new offerings were conducted relying on this amendment within only six months after the exemption became available (Higgins, 2014).

To identify the heterogeneous effect of this regulatory shock across different regions, we measure the size of the accredited investor pool in a region following Lindsey and Stein (2019). An accurate identification of the accredited investor status requires the assessment of both the income and net worth standards, where an accredited investor must have income in excess of $200,000 or a net worth over $1 million. Using the Survey of Income and Program Participation (SIPP), a household-level longitudinal survey covering individuals and families, we compute the number of families qualifying as accredited investors at the state level.[14] The specific topical module of the SIPP we use is the only data source that provides detailed income, assets, and liability information necessary to accurately determine the accreditation status based on both the income and net worth standards. Unfortunately, it does not provide more granular information for regions finer than states, such as MSAs.

[14]We use Wave 10 of the 2008 SIPP, for which the survey was conducted between September and December of 2011.

Therefore, we inevitably use the state-level treatment as our main specification. However, we later conduct robustness tests using MSA-level treatment, relying only on the household-level income data from the American Community Survey to capture MSA-level variation.

We consider the general solicitation shock as an increase in the broad participation of accredited investors, with its effect being stronger in regions where the size of the accredited investor pool is large and thus more likely to be affected by the change. Accordingly, we define the treated and control groups based on the number of qualifying accredited investors prior to the general solicitation shock. Specifically, we estimate the following difference-in-differences regression:

$$Outcome_{m,t} = \alpha + \beta_1 Treated_s \times PostGS_t + \beta_2 Treated_s + \beta_3 PostGS_t + \gamma \Gamma_{m,t-1} + \epsilon_{m,t}. \quad (2)$$

The state-level $Treated$ is an indicator variable that is equal to one if the number of qualifying accredited investors in state s using the pre-shock period income and wealth data is above the sample median and zero otherwise. The MSA-level $Treated$ in our robustness tests is an indicator variable that is equal to one if the percentage of households with income greater than \$200,000 in MSA m using only the pre-shock period income data is above the median and zero otherwise.[15] $PostGS_t$ is one after the shock in 2013 and zero otherwise, and $\Gamma_{m,t-1}$ is a set of MSA-level control variables in the year prior to t, including the number of incumbent firms and other establishments, number of closed firms, number of all types of investors (to control for general investor activity), population (to control for the size of MSA), GDP, median household income, housing price growth, and the presence of the top 50 universities. We include state (or MSA) and year fixed effects and cluster standard errors at the state (or MSA) level consistent with the level of the treatment variable. The state (or MSA) and year fixed effects subsume the standalone $Treated$ and $PostGS$ estimates.

One may be concerned that our treatment regions (states or MSAs) are systematically dif-

[15]For the state-level treatment, we follow the amended definition of accredited investors under the Dodd-Frank Act in 2010, which excludes a person's primary residence from the calculation of net worth. The MSA-level treatment is unaffected by the Dodd-Frank amendment.

ferent from control regions. To address the selection concerns, we further perform propensity score matching (PSM) between the treated and control regions to control for any observable differences between the two groups. For each treatment region, we match two nearest-neighbor control regions on a common support using the propensity score.[16] The propensity score is estimated from a cross-sectional probit model, regressing the treatment indicator on the pre-treatment average values of covariates.[17] The balancing test results of the matching are reported in Table IA.6. In both Panels A and B of Table IA.6, we do find no significant differences in the covariates between the treated and control groups after the matching both for the state and MSA level treatments, ensuring the comparability of the two groups.

To verify whether the general solicitation shock indeed had an economically sizable impact on angel investors' participation, we first consider the total amount of angel funding at the MSA level for $Outcome_{m,t}$. We present the results in the first two columns of Table IA.7. In column (1), the estimate on the interaction term is positive and statistically significant at the 5% level. The magnitude of the estimated coefficient, 6.38, translates into about \$6.4 million higher angel funding in the treated states after the shock. In column (2), we further show the dynamic treatment effects by year. We find that the coefficient estimates for the years before the shock are statistically insignificant while those for the years after the shock are much larger in magnitude and statistically significant. We also provide a visual inspection of the parallel trend assumption by plotting the treatment effects over time in panel (a) of Figure 3. The lack of pre-trends ensures that our results are unlikely to be driven by other observable factors that correlate with the angel funding amount. The effects of the treatment occur with some time lag likely because it was adopted almost at the end of 2013.

Further, we consider analogous tests using the total amount of VC funding as $Outcome_{m,t}$. This is to examine whether the effects of general solicitation are manifested only through accredited angel investors, not others. We present the results in columns (3) and (4) of Table

[16] Our results are robust to using 3 or 5 nearest neighbors.

[17] The covariates for PSM do not include the indicator variable for the presence of the top 50 universities as the variable is a time-invariant MSA-level indicator.

IA.7. We find no significant effects of the general solicitation shock on VC funding. The lack of impact on VC funding is likely because VCs source deals mainly from their networks, other investor referrals, and self-generated pool rather than from solicitations of startups without any connection (Gompers et al., 2020). Hence, the results from these falsification tests ensure that our empirical setting solely exploits the variation in angel investors' participation. We repeat the same exercises using the PSM sample and find our results remain strong and robust in columns (5) through (8) of the table. In the following analyses, we continue to use the general solicitation shock to examine the causal relation between angel investors and the local industry diversity of startups.

5 Business Similarity and Financing

5.1 Business Similarity and Angel Investors

The conceptual framework and the stylized facts that we document in Sections 4.1 and 4.2 predict that the presence of angel investors is positively correlated with startup business diversity as they fund startups from diverse industries. In this section, we empirically test this prediction using the identification strategy that exploits the exogenous changes in angel investors' participation described in Section 4.3. We report the results in Table 3.

[Insert Table 3 Here]

The first three columns use the state-level treatment, which uses more accurate identification of qualifying accredited investors based on both the income and net worth standards, while the last three columns use the MSA-level treatment, relying only on the income standard of accreditation status but fully exploiting the MSA-level variations. All columns use the PSM sample as described in Section 4.3 to mitigate the selection concerns. Each column estimates Equation (2) using the original, vintage, and soft-cosine similarity scores described in Section 3.

In column (1), we find that the business diversity among startup firms as measured by the original similarity score is 11.6% lower in treated states, which is expected to experience a larger increase in angel investors' participation after the general solicitation shock.[18] That is, the increase in the supply of angel funding, which the general solicitation shock aimed at, results in greater local startup diversity. The main result is robust using alternative similarity scores, where the effect is slightly larger when using the vintage score (11.9%) and slightly smaller when using the soft-cosine score (10.7%). The change in local business diversity of about 0.45 in the estimated coefficient is fairly substantial as it accounts for almost 60% of the average decrease in MSA-level business similarity scores over our entire sample period of 0.7. The results are robust using the MSA-level treatment as shown in columns (4) through (6), where the coefficients are larger. The effects range between 20.4% and 30%, capturing greater variation in the MSA-level similarity.[19]

We further show that our results are not driven by any differential trends in local business similarity between the treated and control groups prior to the general solicitation shock. The visual inspection of a parallel trend in Panel (b) of Figure 3, which plots the treatment effects over time, highlights a few important observations. First, the coefficients in the years prior to the treatment are close to zero and statistically insignificant. Hence, it is highly unlikely that our results are driven by unobservable differences between the firms in the treated and control states. The lifting of the prohibition on general solicitation appears to take its full effect in about a year. This is reasonable considering that the provision became effective in September 2013 and thus requires time for firms to put together marketing materials and take necessary steps to verify accredited investor status for participating investors. Consistently thereafter, the treatment effects are negative and statistically significant. We also confirm these results in a regression with the interaction of the treated group and each of the year

[18]For the matching sample, the mean similarity scores ($Similairty \times 100$) in the treated states prior to the shock are 3.854, 3,824, and 4.263 for original, vintage, and soft-cosine scores, respectively.

[19]The mean similarity scores ($Similairty \times 100$) in the treated MSAs prior to the shock are 3.923, 3.911, and 4.355 for original, vintage, and soft-cosine scores, respectively. Again, *Top school* 50 is time-invariant at the MSA-level and is not estimated in the MSA-level regressions with MSA fixed effects. Table IA.8 presents results using the full sample. Results are also consistent without the PSM matching.

dummy variables in Table IA.9. We find robust results across all three different measures of similarity score.

Lastly, we acknowledge that the JOBS Act has three other major titles with different implementation dates.[20] Among the other provisions, Titles 3 and 4, implemented later in May 2016 and June 2015, respectively, could be potentially relevant to angel investors and our empirical setting. Both provisions are grounded on the elimination of the prohibition of general solicitation (Title 2), and their impacts are likely in the same direction as the general solicitation shock, although deemed to be much smaller compared to other exemptions such as Regulation D (McHenry, 2022; Pollner, Wardle, and Hamlette, 2019). In untabulated results, we directly control for the crowdfunding amount (related to Title 3) in the regressions analogous to Table 3 and find our results remain virtually unchanged. However, we are unable to separately control for the Regulation A+ amount (related to Title 4) in the regressions, as it is also included in the amount of angel funding if Regulation A+ capital is funded by angel investors in our sample. Hence, it may slightly amplify the treatment effects around their implementation years, given their possible impacts.

5.2 Business Similarity and Angel Funding

The results in Section 5.1 indicate that angel investor participation has a causal relation with local business diversity. In this section, we directly examine how a startup's business similarity to local peers affects its access to angel financing, considering the portfolio preference of angel investors that we have previously documented.

We regress angel funding on the firm-year level business similarity. For ease of interpretation, we use *dissimilarity*, which is one minus the similarity score and captures a startup

[20]The JOBS Act has four major titles, where the elimination of the prohibition of general solicitation is Title 2. Title 1 aims to reduce public reporting and compliance obligations of small firms, thereby fostering capital raising in the U.S. IPO markets. Title 1 became effective immediately in 2012. Title 3 is the adoption of Regulation Crowdfunding (Regulation CF), aiming to legalize equity crowdfunding by allowing companies to raise small amounts of money from a large number of investors through online platforms. Title 3 became effective in May 2016. Title 4 is the amendment of Regulation A ("Regulation A+"), allowing for larger fundraising efforts by increasing the limit for securities offerings exempted under Regulation A. Title 4 became effective in June 2015.

firm's business difference from its peers within an MSA. Specifically, we estimate the following regression specification:

$$Angel\ Funding_{i,t} = \alpha + \beta Dissimilarity_{i,m,t} + \gamma \Gamma_{i,m,t} + \epsilon_{i,t}, \tag{3}$$

where $Angel\ Funding_{i,t}$ is either an indicator variable for receiving funding 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 score across the pairs of firm i and all incumbent startups in MSA m in year t. The control variables ($\Gamma_{i,m,t}$) include the age of firm i and the number of firms, G Index, and EG Index of MSA m in year t.[21] Firm and year fixed effects are also included, and standard errors are clustered at the MSA level. Table 4 presents the results.

[Insert Table 4 Here]

We find that an increase in business dissimilarity increases a firm's access to angel funding in Panel A. For example, the coefficient estimate for *Dissimilarity* in column (1) implies that one standard deviation (0.031) increase in dissimilarity scores increases the probability of receiving funding from angel investors by 0.47 percentage points, which translates into a 39.3% increase from the unconditional mean of the angel funding likelihood.[22] The magnitudes of the effects are even larger when using the vintage and soft-cosine dissimilarity measures at 80% and 95%, respectively.[23] In column (2), we find that a one standard deviation increase in dissimilarity score increases the amount of angel funding by 37.5%.

[21] 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.

[22] This 39.3% is computed $0.031 \times 0.152/0.012$, where 0.012 is the unconditional mean of the angel funding likelihood.

[23] Although we show the results using the vintage similarity for the robustness of our finding, the implications of the original similarity score are more economically relevant as it conveys that angel investors are more willing to fund startups with emerging businesses such as Artificial Intelligence and Cryptocurrencies that did not exist at the beginning of our sample period.

The results in this table capture the extensive margin of the effect. For the intensive margin of the effect, we only consider the funding sample in Panel B of Table 4. Only 15% of our sample firms receive any funding over our sample period. We find the results are consistent and larger for the funding sample, where firms in the sample have received at least one funding during the sample period. A one standard deviation (0.028) increase in dissimilarity scores in the sample increases the likelihood of receiving angel funding and the funding amount by 19 percentage points (70%) and $233K (46%), respectively. These results collectively show the micro-foundation of our finding in the previous section that the increase in local startup diversity following the general solicitation shock is indeed driven by angel investors' preference for funding dissimilar startups in the area.

5.3 Alternative Explanations and Robustness

Although angel funding is considered the most feasible access to capital for very early-stage startups, some startups receive VC funding from the outset. Hence, one may be concerned if our results hold for all types of entrepreneurial funding regardless of the portfolio preferences of angel investors. If our angel funding results are indeed driven by angel investors' preference for firms geographically close but diverse in industry, we should not find similar results for VC funding. In Table IA.10, we examine this possibility using VC funding as the dependent variable. We find that the effects of business dissimilarity on VC funding are exactly the opposite. The results show that a one standard deviation increase in dissimilarity scores decreases the probability of receiving VC funding by 0.4 and 6 percentage points, respectively in the full sample of column (1) and the funding sample of column (3). Similarly, a one standard deviation increase in dissimilarity decreases the amount of VC funding by $88,000 and $5.5 million in columns (2) and (4), respectively.

These results are critical for the interpretation of our main results for several reasons. First, results in Table IA.10 help clearly identify angel investors' portfolio preferences as the key underlying mechanism of our findings in Tables 3 and 4. Second, these results also help

mitigate the concern that an unobserved local economic shock drives our results. If some regional economic factor were attracting many new startups (thereby raising dissimilarity) and simultaneously increasing the local funding opportunities, we would observe the same directional effects for VC funding, which we do not.

To address any potential concerns related to why a startup is founded in a specific location, such as the influence of top colleagues and local startup incubators, we exclude all entry-year observations of startups in our sample. This approach forces the variation in a firm's dissimilarity scores to be only due to subsequent external disturbances by the entry or exit of other firms in the area.[24] We present the results in Table 5 and find that our results remain robust and only become stronger.

[Insert Table 5 Here]

Next, we show in a series of robustness tests that our results are not spuriously driven. First, one may be concerned that Crunchbase funding data is incomplete. To address the concern, we follow Denes et al. (2020) and supplement angel investment data with the SEC Form-D filings. We only include observations for private equity offerings from the Form-D data by aggregating the total amount raised in each filing. We find that our results remain robust in Panel A of Table IA.11. Second, one may be concerned that small, individual angel investors tend to selectively report only successful investments and the self-reporting feature of Crunchbase data may drive our results. To address this concern, we use the time stamps of investment reporting and exclude any reports made more than two years after the funding announcement. Angel investors typically hold their portfolio firms for four to six years (MIT Entrepreneurship Center, 2000). Thus, reporting their investment within two years after the investment is made is likely timely, mitigating the self-selection bias. In Panel B of Table IA.11, our results still remain strong and consistent albeit with slightly weaker magnitudes.

Lastly, we show that our results are not driven by the correlation between dissimilarity

[24]The magnitudes of the post-entry variations in similarity scores from the entry or exit of other firms in the same MSA is on average 5.7% (0.0021) from the mean similarity score of (0.0376), large enough to lead to meaningful external disturbances to the dissimilarity scores.

scores and startup quality. For example, one may be concerned that a low average similarity score in an MSA indicates low local competition and, thus, a lack of high-quality firms that would have been chosen to receive VC funding. We rule out this possibility in Table IA.12. We consider a regression model of the Entrepreneurship Quality Index (EQI) on MSA-level dissimilarity scores. We find that MSA-level dissimilarity scores are not related to regional entrepreneurship quality, even after controlling for various MSA-level economic factors and other entrepreneurial indices (entrepreneurship cohort potential index and regional ecosystem acceleration index). This result indicates that our findings are not driven by low-quality firms left to receive angel funding because they were unable to receive VC funding.

5.4 Fostering Emerging Industries and Individual Diversity

We further focus on providing the details on how angel investors attain the industry diversity of their portfolios. First, we explore the types of diverse startups that angel investors prefer. Business diversity can be achieved by adding firms from two very different ends of the technology spectrum, such as funding a distillery versus a firm developing artificial intelligence technology. The economic implications of funding one type of diverse startup over the other can vary significantly. We explore this question in Table 6.

[Insert Table 6 Here]

We run tests analogous to the main specification in Table 4 but add an interaction term between dissimilarity scores and an indicator for emerging industries during our sample period. *Emerging* is equal to one if the funded firm belongs to any of the 34 emerging industries identified from our list of Crunchbase categories, such as Artificial Intelligence, Cloud Computing, Machine Learning, and Virtual Reality.[25] We find that angel investors

[25]We prompted ChatGPT to identify emerging industries among our 742 Crunchbase categories, specifically relevant to our sample period. It selected 34 of them as emerging industries based on the following criteria: technological advancement, changing consumer trends, market disruption, environmental sustainability, and digital transformation. We present the list in Table IA.13. Many of these categories overlap with the newly trending categories shown in Table IA.2.

strongly prefer to fund startups in these emerging industries. The coefficients on the interaction term are consistently positive and significant across all columns with different measures of similarity scores while the coefficients on the standalone measure of dissimilarity also remain significantly positive. These results have an important implication that angel investors play a significant role in providing essential funding to support emerging businesses around their inception, thereby fostering innovation in emerging markets.

Next, we examine whether the characteristics of individual angel investors in local areas contribute to the positive relationship between the industry diversity of startups and angel funding. Angel investors typically gather information within the areas where they live to identify investment opportunities. Moreover, they often engage in local angel groups to enhance their investment strategies by learning from other group members (Cable, 2011). We are particularly interested in knowing how the diversity among local angel investors possibly influences their investment decisions to fund diverse firms and, as a result, marginally increases a dissimilar startup's chance of receiving angel funding. We specifically consider two dimensions of individual angel investor characteristics: demographics and education.

In Table 7, we repeat the analyses in Table 4 but additionally interact dissimilarity scores with an indicator for local angel diversity. *High Diversity* is an MSA-year level variable that equals one if the diversity measure based on either demographics or education of angel investors is above the median and zero otherwise. The diversity measure is computed as the sum of standardized 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 of Table 7 report the results based on demographics and education, respectively.

[Insert Table 7 Here]

Columns (1) and (2) in Panel A show that startup firms with higher dissimilarity scores have a significantly higher likelihood and amount of angel funding when the region exhibits

greater diversity among angel investors in their demographic characteristics. The incremental effect from *High Diversity* is approximately 8% of the primary effect of the dissimilarity scores. This finding is robust to using other similarity measures. The results in Panel B based on educational diversity are broadly consistent with those in Panel A and indicate that dissimilar startup firms are marginally more likely to secure angel funding when the region has more diverse angel investors in their educational backgrounds.

The findings in this section are especially noteworthy. First, the results convey that angel investors promote the business diversity of startups through their willingness to explore and provide capital to firms especially in new and emerging industries at the forefront of technology. These firms face greater capital market frictions due to relatively high uncertainties in their ultimate success. Angel investors' exploration of such projects may stem from their smaller funding size and limited involvement with funded firms, mitigating the downside of their investments. Among the startups funded by angel investors, promising ones would survive and eventually get funded by later-stage financiers such as VCs. Second, the diversity of local individual investors is economically conducive to the entrepreneurial ecosystem. Angel investors' geographic constraints naturally create local clustering of angel investors with diverse human capital and backgrounds. The diversity of these local groups thus significantly influences sourcing and evaluating local deals as they arise. These findings carry important policy implications suggesting that fostering diversity among individual investors in a region can be a driving force in nurturing unique business ideas among startups.

6 Broad Implications of Angel Investment Strategies

Thus far, we have shown distinct and consistent funding patterns among angel investors. However, if these funding patterns simply reflect a lack of investment sophistication, our findings would suggest a misallocation of valuable capital that could have been directed toward more promising startups. Therefore, in this section, we examine the outcomes of

angel investment based on their unique funding preferences. We also aim to determine whether these funding patterns are meaningful and impactful for firms, thus influencing startups' entry decisions.

6.1 Angel Portfolio Outcomes

We first examine the portfolio outcomes of angel investors. While we cannot explicitly test for an optimal funding strategy at each round under our bandit model framework, we can at least assess whether their strategies to fund diverse startups are consistent with generating positive net present value (NPV) at each investor level. To do so, we consider several measures of success of funded startups by estimating the following regression specification:

$$Outcome_j = \alpha + \beta_1 Angel\ Investor_j + \beta_2 Dissimilar_j + \beta_3 Angel\ Investor_j \times Dissimilar_j + \gamma \Gamma_j + \epsilon_j, \tag{4}$$

where $Outcome_j$ is one of the dependent variables, including the likelihood of reaching subsequent funding rounds and achieving a successful exit through IPO or acquisition, among all funded firms by investor j. $Dissimilar_j$ is an indicator for whether investor j's portfolio is identified as following the diverse-industry strategy—defined as having an overall similarity score among funded firms in the investor's portfolio below the median of all portfolios of investors of the same type. Our main variable of interest is the interaction between $Dissimilar_j$ and $Angel\ Investor_j$ to examine whether the diverse-industry strategy is specifically relevant to angel investments. Γ_j is a set of investor-level control variables, including the number of total investments, the investing years since investor j made her first investment, and the average funding amount. Table 8 presents the results.

[Insert Table 8 Here]

Column (1) reports the results of subsequent funding rounds. The positive coefficient on *Angel Investor* suggests that angel funding likely precedes VC funding. More importantly, we find that the interaction term of *Angel Investor* 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. The coefficient translates into a 13.6 percentage point increase, or a 20% increase from the unconditional mean, in the likelihood of reaching subsequent funding rounds. In columns (2) and (3), we examine the fraction of firms with eventual IPO and/or acquisition exits in the investor's portfolio. We first note that angel investors generally have significantly lower exit rates relative to VCs based on the significantly negative coefficient on *Angel Investor*. However, the diverse-industry strategy appears to significantly improve the overall IPO and/or acquisition exit rates of an angel investor's portfolio. The estimated effects are 3.4 and 7 percentage point increases in the portfolio's IPO exit rate (column 2) and IPO and acquisition exits together (column 3), respectively. These results translate into as large as a 32% increase in the successful exit rate from the unconditional mean.[26]

6.2 Startup Entry Decisions

Lastly, we explore whether angel investors' unique funding preferences have any impact on startups' entry decisions. We hypothesize that startup firms aiming to receive angel funding will consider the preferences of angel investors when founding their business and choosing a location where their business appears to be more diverse compared to incumbent startups. To explore startup entry decisions, we create firm-level cross-sectional data at the time of the entry year t and compute a firm's similarity to the incumbent startups in each of the 416 MSAs in the year prior to the entry. For each firm in the sample, this process generates one actual similarity score in the chosen MSA and 415 counterfactual similarity scores in unchosen MSAs. We run firm-MSA-entry year level regressions using the DID design similar to Table 3 and additionally interact the DID estimates with an entering firm's dissimilarity scores. Note that this regression compares different sets of entering firms before and after the

[26]There may be a concern that investors with only one investment may drive our results. However, our similarity score can only be computed when each investor makes more than one investment, and thus the case falls into $Dissimilar = 0$. We also find consistent results after excluding investors with only one investment.

general solicitation shock since each firm has only one entry year. We specifically estimate the following regression specification:

$$\begin{aligned} Entry_{i,m} = \alpha + \beta_1 Dissimilarity_{i,m} \times Treated_s \times PostGS_t + \beta_2 Treated_s \times PostGS_t \\ + \beta_2 Dissimilarity_{i,m} \times Treated_s + \beta_3 Dissimilarity_{i,m} \times PostGS_t \\ + Dissimilarity_{i,m} + Treated_s + PostGS_t + \gamma \Gamma_m + \epsilon_{i,m}. \quad (5) \end{aligned}$$

The dependent variable, $Entry_{i,m}$, is one if MSA m is chosen by entering firm i and zero otherwise. We then compute firm i's business similarity scores to all potential incumbent startups in MSA m in the year prior to the firm's entry. 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 startup j in MSA m. We denote the average business similarity for firm i in MSA m with respect to all incumbent startups in the MSA by $Similarity_{i,m}$. For ease of interpretation, we use *Dissimilarity* by subtracting *Similarity* from one. Γ_m is a set of MSA-level control variables in the year prior to the entry year, including the number of incumbent startup firms, number of closed firms, number of investors, GDP, net jobs created, housing price index, and presence of top-ranking college. We include state, firm, and entry-year fixed effects. Standard errors are clustered at the firm level since the decision is made at the firm level. The results are presented in Table 9. To conserve space, we report only the estimated coefficients for $Dissimilarity \times Treated \times PostGS$ (the main variable of interest), $Treated \times PostGS$, and *Dissimilarity* along with the control variables.

[Insert Table 9 Here]

Column (1) shows the entry decision using the MSA-level similarity score. Generally, firms seem to weigh in on entering areas with more similar incumbents, partly due to potential agglomeration benefits. This is captured in the negative coefficient on *Dissimilarity*. However, the decision to enter an MSA becomes more likely when being dissimilar is ad-

vantageous for obtaining angel funding as shown by the positive and statistically significant coefficient on the triple interaction term. In other words, disparate startup firms prefer to launch their business in an MSA with more angel participation. This effect is economically significant. A one standard deviation increase in dissimilarity (0.037) is associated with a 3.4% increase from the unconditional probability of entering an MSA randomly.[27] The coefficient on *Top college* 50 is positive and significant in columns (1) and (2), suggesting that the presence of top colleges in the area is also an important consideration at the broader MSA and CSA level, possibly through entrepreneurship programs offered by institutions. Nevertheless, startups appear to strongly account for the angel investors and their preference for disparate startups even after controlling for the college effect. On the other hand, we find that startups tend to avoid highly competitive areas, as reflected in the negative coefficients on the number of closed firms and net jobs created.

In columns (2) and (3), we consider alternative spatial units for entry decisions and the similarity score to understand where the impact of angel investors might be largest and how far the spillovers can reach for startups' entry decisions. We recompute the alternative similarity scores using incumbent startups at each Combined Statistical Area (CSA) and county level, respectively.[28] We find that our results are robust to using broader (CSA) and narrower (county) spatial units with economic magnitudes of the triple interaction terms. The marginal effect of dissimilarity on startups' entry decisions is strongest at the CSA level, as founders typically consider broader regional economy and market opportunities. This effect is also strong at the MSA level, which is congruent with the spatial boundary of angel investors' investment decisions. The county-level effect is weaker, which is not surprising given that the average size of a county is only about 1,200 square miles or approximately 20

[27]The probability of a random choice of one MSA out of the total 416 MSAs is 0.0024. The standard deviation of *Dissimilarity* in this analysis, 0.037, is different from 0.031 from our primary sample because we also consider counterfactual similarity scores of a firm to the incumbent firms in unchosen MSA locations for this analysis.

[28]There are a total of 1,983 counties in the data, resulting in roughly 162 million observations (1,983×81,718) with 81,718 entering firms. For feasibility reasons, we use randomly selected 15% of entering firms in our sample for the county analysis. This approach leaves us with a manageable data size that is comparable to the data size used in our main analyses.

miles in radius—a size too restricted for a location decision.

Overall, this section highlights two important implications regarding angel investors and their funding of startups. First, despite their relative lack of experience compared to VCs, angel investors still manage to evaluate promising startups across diverse industries and provide crucial funding to support firm survival through their early stages. This finding aligns with Kerr et al. (2014) who find that angel investors improve survival rates, exits, and various firm outcomes. Second, angel investors appear to play a significant role in the recent trend in geographically dispersed formations of startups, especially in high-tech industries, in smaller cities across the United States. The high living costs and severe competition in Silicon Valley are partly responsible for this recent geographic dispersion of startups (The Economist, 2018). Other key factors include difficulties in accessing VC financing[29] and technological advances.[30] These changes collectively diminish startups' reliance on industry agglomerations and elevate the relative benefits of launching businesses in areas where other types of early-stage financing, particularly angel funding, are more accessible.

7 Conclusion

In this paper, we show a previously unexplored link between the industry composition of local startups and angel financing. In doing so, we introduce a new measure for startup business similarity using detailed industry classifications available from 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 first document angel investors' unique portfolio preference for funding startups from diverse industries that are close in geographic distance. Angel investors' portfolio preferences distinctively contrast those of VCs, who prefer specialization in a small number of industries,

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

[30]For example, Silicon Valley's own products and services, such as cloud computing, video-conferencing, and online collaboration, have enabled knowledge and labor spillovers in various locations.

and closely resemble the reinforcement learning observed in a bandit model. Based on the angel investor's preference for diverse startups, we show that angel investors play a critical role in shaping a diverse startup ecosystem through their funding of disparate startups, especially from emerging industries at their inception, serving as a vital channel for fostering innovation in the economy.

Furthermore, we find that the demographic and educational diversity among individual angel investors in a region reinforces the positive relationship between business diversity and angel financing. Our findings suggest that policies designed to increase the number of angel investors and enhance their demographic or educational diversity can nurture the development of new and unique business ideas. Finally, we show that angel investors' strategies to fund diverse startups are consistent with generating positive investment outcomes at the investor level. Further, startup firms aiming to receive angel funding consider such strategies of angel investors when founding their business and choosing a location. Taken together, our results offer new insight into the specific mechanism driving the new trend of geographic dispersion with industry diversity in firm formation.

References

Agrawal, Shipra, and Navin Goyal, 2012, Analysis of thompson sampling for the multi-armed bandit problem, in *Conference on learning theory*, 39–1, JMLR Workshop and Conference Proceedings.

Andrews, Raymond, Catherine Fazio, Jorge Guzman, Yupeng Liu, and Scott Stern, 2022, The startup cartography project: Measuring and mapping entrepreneurial ecosystems Research Policy.

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

Auer, Peter, Nicolo Cesa-Bianchi, and Paul Fischer, 2002, Finite-time analysis of the multi-armed bandit problem, *Machine learning* 47, 235–256.

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

Bryan, Kevin, and Jorge Guzman, 2023, Entrepreneurial migration Review of Economics and Statistics.

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, 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, 2010, Buy local? The geography of successful and unsuccessful venture capital expansion, *Journal of Urban Economics* 67, 90–102.

Chen, Jun, and Michael Ewens, 2023, Venture capital and startup agglomeration, *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.

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

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* 141, 454–478.

Hellmann, Thomas F, and Veikko Thiele, 2019, Fostering entrepreneurship: Promoting founding or funding?, *Management Science* 2019, 2502–2521.

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.

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, 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.

Manso, Gustavo, 2011, Motivating innovation, *The Journal of Finance* 66, 1823–1860.

March, James G, 1991, Exploration and exploitation in organizational learning, *Organization science* 2, 71–87.

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

McHenry, Patrick, 2022, Ten years of jumpstart our business startups (jobs) act of 2012, *Committee on Financial Services of U.S. House of Representatives* .

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

Moscarini, Giuseppe, and Lones Smith, 2001, The optimal level of experimentation, *Econometrica* 69, 1629–1644.

Pollner, Glenn, Peter Wardle, and Thurston Hamlette, 2019, The expansion of regulation a. Harvard Law School Forum on Corporate Governance. https://corpgov.law.harvard.edu/2019/01/23/the-expansion-of-regulation-a/ .

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.

Roberts, Kevin, and Martin L Weitzman, 1981, Funding criteria for research, development, and exploration projects, *Econometrica* 1261–1288.

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.

Thompson, William R, 1933, On the likelihood that one unknown probability exceeds another in view of the evidence of two samples, *Biometrika* 25, 285–294.

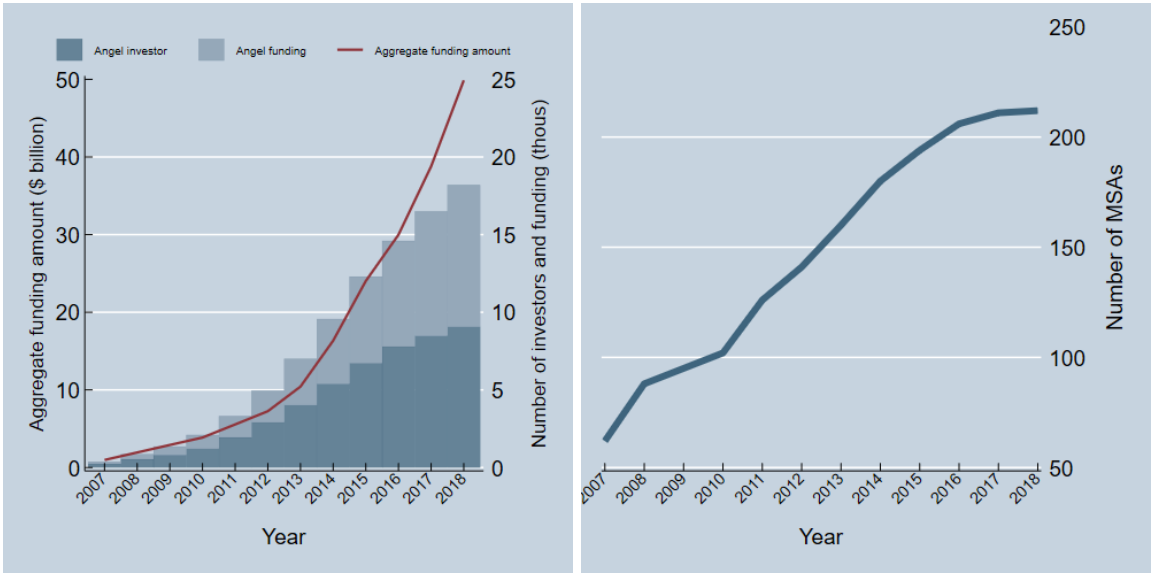
Wang, Xinxin, 2018, Catering innovation: Entrepreneurship and the acquisition market Working Paper.

Watkins, Christopher John Cornish Hellaby, 1989, Learning from delayed rewards .

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

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

The figures show the growth of angel investors and their funding amount from Crunchbase over our sample period between 2007 and 2018. In Panel (a), the red line shows the cumulative aggregate amount of angel funding in $ billion (left axis). The darker blue bars represent the cumulative number of unique angel investors, and the lighter blue bars represent the cumulative number of angel funding. In Panel (b), we show the number of MSAs with at least one angel investor present. The total number of MSAs is 416.

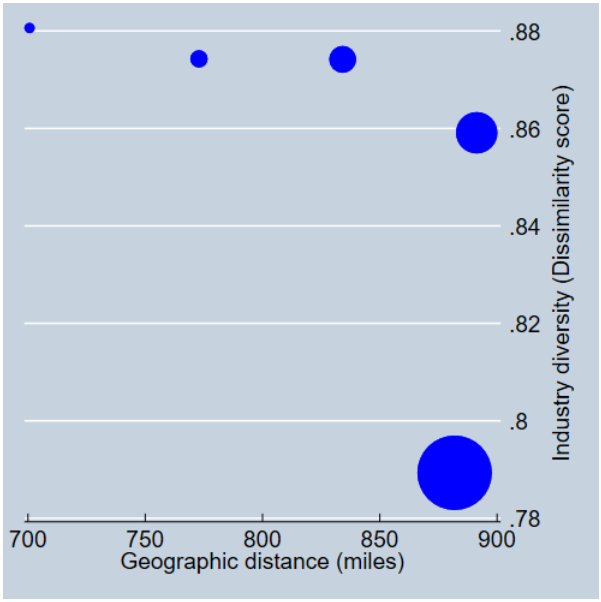


(a) Cumulative funding amount and number of angels and funding

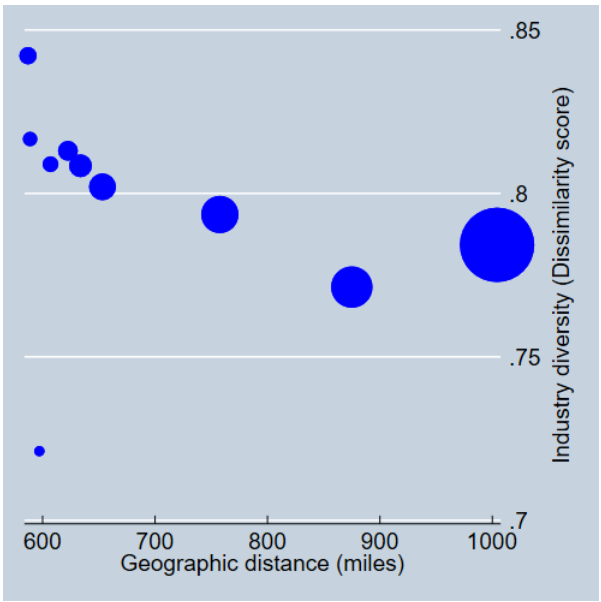
(b) Number of MSAs with angel investor presence

Figure 2: Heterogeneity in Angel Investment Strategies by Investing Experience

The figure shows the angel investor investment strategies by investing experience. *Industry diversity* measures the business dissimilarity among an angel's portfolio firms. *Geographic distance* measures the average distance between an angel and its portfolio firms. In Panel (a), the size of the markers denotes the sequence of investing experiences, where the marker becomes larger as the number of investing experiences increases. The five markers each comprise each angel investor's 1st-2nd, 3rd-4th, 5th-7th, 8th-10th, and 10th-20th investment observations in sequence. In Panel (b), the size of the markers denotes the size deciles of angel investment, where the marker becomes larger as the investment size grows. Each marker represents the average geographic distance and industry diversity of all funding observations in a given bin. For variable definitions and further details of their construction, see Internet Appendix B.



(a) By investment sequence

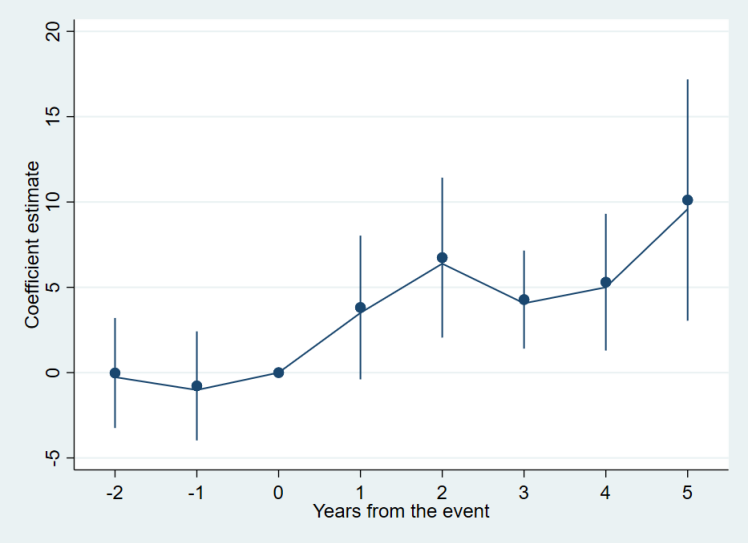


(b) By investment size

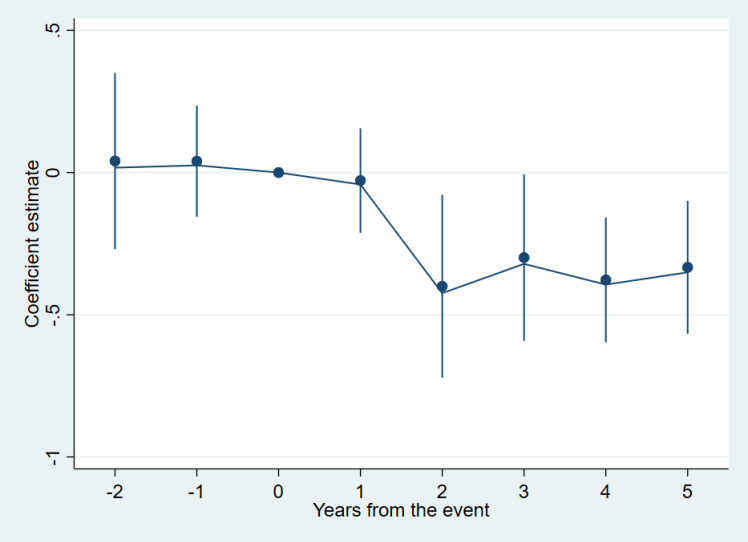
Figure 3: The Effect of General Solicitation on Angel Funding and Industry Diversity

The figures show the treatment effects of the general solicitation shock in 2013 on MSA-level total angel funding and business similarity using the propensity score matching sample. The sample excludes years before 2011 to mitigate any confounding effects of the 2010 Dodd-Frank amendment of the accredited investor eligibility. The y-axis indicates the treatment effect coefficients from the following specification. The outcome variable is the MSA-level aggregate angel funding in Panel (a) and the MSA-level business similarity score in Panel (b). Error bars show the 90% confidence bounds.

$$Outcome_{m,t} = \alpha + \beta_1 \sum_{n=-2}^{5} (Treated_m \times GS_{t+n} + \beta_2 Treated_m + \beta_3 GS_{t+n}) + \beta_4 \Gamma_{m,t-1} + \epsilon_{m,t}$$



(a) MSA-level total angel funding



(b) MSA-level business similarity

Table 1: Rise of New Startup Hubs

The table presents the top 30 Metropolitan Statistical Areas (MSAs) with the highest angel investor growths over the 2007-2018 sample period. *Angel* is the growth rate in the number of angel investors between 2007 and 2018 in percentage. *Similarity* is the growth rate of the average business similarity score of an MSA (described in Section 3) between 2007 and 2018 in percentage. *Firm* is the growth rate in the number of startups in an MSA between 2007 and 2018 in percentage. San Francisco and San Jose, which are considered as the Silicon Valley, are denoted with ⋆. MSAs that are often mentioned as new rising startup hubs in the media are denoted with †, which come from media sources such as Bloomberg, the Economist, and Kenan Insights report as referenced in the footnote 2. The average growth rates of angel investors, business similarity, and the number of firms across the top 30 MSAs are reported at the bottom and compared to the all MSA average. *** indicate the statistical significance of the difference between the top 30 MSA and all MSA averages at the 1% level.

	Growth rates		
	(1)	(2)	(3)
MSA Name	Angel	Similarity	Firm
Nashville-Davidson-Murfreesboro-Franklin, TN	35.0	-22.2	52.1
Phoenix-Mesa-Scottsdale, AZ	27.5	-6.7	46.5
Detroit-Warren-Livonia, MI	22.0	-26.3	24.6
Kansas City, MO-KS	22.0	-12.2	41.9
St. Louis, MO-IL	17.5	-7.2	34.6
† Chicago-Naperville-Joliet, IL-IN-WI	13.7	-15.5	51.7
† Portland-Vancouver-Beaverton, OR-WA	13.7	-16.8	49.6
Sacramento-Arden-Arcade-Roseville, CA	13.0	0.4	35.9
† New York-Northern New Jersey-Long Island, NY-NJ-PA	10.7	-3.4	86.0
Miami-Fort Lauderdale-Pompano Beach, FL	10.7	-5.3	79.0
† Austin-Round Rock, TX	10.2	-15.8	88.7
Richmond, VA	10.0	5.9	34.1
Atlanta-Sandy Springs-Marietta, GA	10.0	-13.7	40.1
Bridgeport-Stamford-Norwalk, CT	9.7	-18.8	34.1
† Boston-Cambridge-Quincy, MA-NH	9.7	-12.4	48.6
Cleveland-Elyria-Mentor, OH	9.5	-0.3	29.8
Tulsa, OK	9.0	-25.9	18.0
New Orleans-Metairie-Kenner, LA	9.0	-9.1	63.1
Washington-Arlington-Alexandria, DC-VA-MD-WV	9.0	-29.8	37.2
Baltimore-Towson, MD	8.5	-11.2	41.4
† Los Angeles-Long Beach-Santa Ana, CA	8.5	-2.7	77.7
Las Vegas-Paradise, NV	8.2	-4.9	105.9
Rochester, NY	8.0	1.8	18.5
† Denver-Aurora, CO	7.9	-11.5	58.2
Raleigh-Cary, NC	7.8	-3.8	30.2
⋆ San Francisco-Oakland-Fremont, CA	7.6	-3.8	117.5
† San Diego-Carlsbad-San Marcos, CA	7.4	-14.2	61.0
Charlotte-Gastonia-Concord, NC-SC	7.3	-18.2	46.0
Houston-Sugar Land-Baytown, TX	7.3	-9.6	31.6
† Seattle-Tacoma-Bellevue, WA	7.2	-3.0	66.7
† Durham, NC	7.0	-31.6	61.1
⋆ San Jose-Sunnyvale-Santa Clara, CA	6.2	-18.6	84.2
Averages of the above MSAs (with the highest angel growth rates)	11.5***	-11.4***	52.9***
Averages of all MSAs	5.92	-0.88	33.57

Table 2: Summary Statistics of Crunchbase Variables

The table provides summary statistics on firm age, funding information, and similarity scores. The full sample comprises 117,944 unique firms, whereas the funding sample comprises 18,436 unique firms that reported at least one funding round during our sample period between 2007 and 2018. Panel A describes firm ages and funding information. The top panel presents firm-level summary statistics. The middle and bottom panels present funding-level descriptive statistics by investor types. There are 47,083 funding round observations. Panel B describes the firm-year-level business similarity scores using the full sample. For variable definitions and further details of their construction, see Internet Appendix B.

Panel A: Firm age and funding information (using funding sample)

	Mean	Sd	Min	p50	Max
Firm-level					
All firm age (full sample)	4.33	3.04	0	4	10
Funded firm age	3.81	2.83	0	3	10
Number of funding rounds	1.90	1.33	1	1	29
Age at funding	3.07	2.48	0	3	10
Age at first funding	2.48	2.48	0	2	10
Funding-level: Angel					
Angel (dummy)	0.19	0.39	0	0	1
Funded firm age	2.35	2.13	0	2	10
Funded firm age at first funding	1.75	1.91	0	1	10
Number of investors	2.07	2.14	1	1	61
Funding amount ('000s)	2,425.00	9,468.54	0	875	499,505
Distance to investors (miles)	549.89	791.22	0	88	3,294
Funding-level: VC					
VC (dummy)	0.59	0.49	0	1	1
Funded firm age	3.38	2.59	0	3	10
Funded firm age at first funding	2.52	2.46	0	2	10
Number of investors	2.41	1.74	1	2	31
Funding amount ('000s)	11,460.58	46,001.16	1	4,286	4,620,000
Distance to investors (miles)	724.01	809.39	0	427	4,972

Panel B: Business similarity score (using full sample)

	Mean	Sd	Min	p50	Max
Similarity	0.0376	0.0305	0.000	0.0309	0.6035
Vintage similarity	0.0384	0.0311	0.000	0.0314	0.4082
Soft-cosine similarity	0.1208	0.0724	0.000	0.1088	0.5697

Table 3: Angel Investors and Local Startup Similarity

The table shows results from the regressions that examine the effects of the general solicitation shock on local startup industry diversity using the MSA-year level observations using the propensity score matching sample. We use the elimination of the general solicitation prohibition of Rule 506(c) as an exogenous shock to angel investors' participation. *Similarity* is the business similarity score. The state-level treatment sample excludes years before 2011 to mitigate any confounding effects of the 2010 Dodd-Frank amendment of the accredited investor eligibility. The state-level *Treated* is an indicator variable that is equal to one if the number of accredited investors during the pre-shock period is above the median and zero otherwise. The MSA-level treatment sample is unaffected by the Dodd-Frank amendment and uses full sample period between 2007 and 2018. The MSA-level *Treated* is an indicator variable that is equal to one if the percentage of households with income greater than $200,000 during the pre-shock period is above the medium and zero otherwise. *PostGS* is an indicator variable that is equal to one after the general solicitation shock in 2013 and zero otherwise. Control variables are lagged by one year. Standard errors are clustered at the state level. For variable definitions and further details of their construction, see Internet Appendix B.

	Similarity × 100					
	State-level treatment			MSA-level treatment		
	Original	Vintage	Soft-cosine	Original	Vintage	Soft-cosine
	(1)	(2)	(3)	(4)	(5)	(6)
Treated×PostGS	-0.448**	-0.454**	-0.454**	-1.175**	-0.959**	-0.891**
	(0.176)	(0.173)	(0.169)	(0.469)	(0.446)	(0.449)
Ln(Firms)	0.469**	0.357	0.374	0.148	0.181	0.201
	(0.180)	(0.233)	(0.277)	(0.364)	(0.355)	(0.353)
Ln(Firm Deaths)	0.969	1.418*	1.648*	-0.197	0.206	0.153
	(0.914)	(0.769)	(0.824)	(0.445)	(0.523)	(0.509)
Ln(Investors)	-0.136	0.044	0.087	0.121	0.080	0.100
	(0.137)	(0.127)	(0.154)	(0.153)	(0.170)	(0.165)
Ln(Population)	-2.992***	-3.026**	-3.027**	6.305*	4.675	4.396
	(1.083)	(1.156)	(1.269)	(3.634)	(3.246)	(3.337)
Ln(GDP)	2.300**	1.962**	2.154***	-0.342	-0.800	-0.605
	(0.930)	(0.778)	(0.793)	(0.711)	(0.773)	(0.777)
Ln(Establishments)	-1.140	-1.371	-1.890	-4.467	-3.071	-3.128
	(1.256)	(1.168)	(1.187)	(2.900)	(2.628)	(2.698)
Ln(MedHHIncome)	0.492	0.791	0.933	1.278	2.135	2.141
	(1.121)	(1.097)	(1.180)	(1.287)	(1.951)	(1.952)
HPI growth	1.418	1.238	1.525	0.278	0.439	0.410
	(1.149)	(1.038)	(0.954)	(0.583)	(0.502)	(0.504)
Top college 50	-0.817	-0.683	-0.631			
	(0.550)	(0.537)	(0.542)			
State FE	Y	Y	Y	N	N	N
MSA FE	N	N	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	2325	2325	2325	2284	2284	2284
Adjusted R^2	0.535	0.468	0.512	0.745	0.688	0.711

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

Table 4: Business Dissimilarity and Angel Funding

The table shows results from the firm-year level regressions of angle funding on dissimilarity. Panel A uses the full sample, and Panel B uses the funding sample, which only includes firms that have at least one funding record in our sample. *Dissimilarity* is one minus the similarity score. In each set of two columns, *Dissimilarity* is measured using the original, vintage, and soft-cosine versions as described in Section 3. $\mathbb{1}(Angel)$ 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 B.

Panel A: Full sample

	Original		Vintage		Soft-cosine	
	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)
Dissimilarity	0.152***	0.033***	0.298***	0.066**	0.347***	0.077***
	(0.054)	(0.012)	(0.115)	(0.027)	(0.124)	(0.029)
Ln(Age)	0.002**	0.001***	0.002**	0.001***	0.002**	0.001***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Ln(Firms)	-0.040***	-0.010***	-0.041***	-0.010***	-0.040***	-0.010***
	(0.010)	(0.002)	(0.010)	(0.002)	(0.010)	(0.002)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	728332	728332	725435	725435	728332	728332
Adjusted R^2	0.135	0.133	0.135	0.133	0.135	0.133

Panel B: Funding sample

	Original		Vintage		Soft-cosine	
	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)
Dissimilarity	6.776***	8.357***	6.947***	8.379**	6.730***	9.357***
	(1.058)	(2.903)	(1.151)	(3.416)	(1.097)	(3.266)
Ln(Age)	-0.074***	0.351***	-0.075***	0.350***	-0.075***	0.351***
	(0.021)	(0.101)	(0.020)	(0.102)	(0.021)	(0.101)
Ln(Firms)	-0.616***	-0.257	-0.622***	-0.265	-0.602***	-0.232
	(0.098)	(0.249)	(0.100)	(0.244)	(0.101)	(0.248)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	21151	21151	21133	21133	21151	21151
Adjusted R^2	0.372	0.249	0.372	0.249	0.372	0.249

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

Table 5: Business Dissimilarity and Angel Funding – Dropping entry-year observations

The table shows results from the firm-year level regressions of angel funding on dissimilarity. We drop the entry-year observations to mitigate selection concerns. In each set of two columns, *Dissimilarity* is measured using the original, vintage, and soft-cosine versions as described in Section 3. $\mathbb{1}(Angel)$ 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 B.

	Original		Vintage		Soft-cosine	
	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)
Dissimilarity	0.164**	0.038**	0.345***	0.081***	0.405***	0.095***
	(0.066)	(0.015)	(0.131)	(0.031)	(0.144)	(0.034)
Ln(Age)	-0.021***	-0.005***	-0.021***	-0.005***	-0.021***	-0.005***
	(0.007)	(0.002)	(0.007)	(0.002)	(0.007)	(0.002)
Ln(Firms)	-0.045***	-0.011***	-0.045***	-0.011***	-0.044***	-0.011***
	(0.013)	(0.003)	(0.013)	(0.003)	(0.012)	(0.003)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	644680	644680	642139	642139	644680	644680
Adjusted R^2	0.155	0.153	0.155	0.153	0.155	0.153

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

Table 6: Business Disimilarity and Angel Funding in Emerging Industries

The table shows results from the firm-year level regressions of angel funding on dissimilarity. *Dissimilarity* is one minus the similarity score. In each set of two columns, *Dissimilarity* is measured using the original, vintage, and soft-cosine versions as described in Section 3. *Emerging* is an indicator variable that is equal to one if a startup is in one of the emerging industries and zero otherwise. Emerging industries are characterized by advancements in technology, shifts in consumer behavior, or changes in market dynamics. $\mathbb{1}(Angel)$ is an indicator 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 B.

	Original		Vintage		Soft-cosine	
	$\mathbb{1}$(Angel)	\$Funding	$\mathbb{1}$(Angel)	\$Funding	$\mathbb{1}$(Angel)	\$Funding
	(1)	(2)	(3)	(4)	(5)	(6)
Dissimilarity×Emerging	1.093***	0.269***	1.446***	0.344***	1.803***	0.419***
	(0.315)	(0.077)	(0.399)	(0.092)	(0.451)	(0.107)
Dissimilarity	0.118**	0.024**	0.249**	0.054**	0.282***	0.062***
	(0.049)	(0.011)	(0.102)	(0.023)	(0.104)	(0.024)
Ln(Age)	0.002**	0.001***	0.002**	0.001***	0.002**	0.001***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Ln(Firms)	-0.040***	-0.010***	-0.041***	-0.010***	-0.039***	-0.010***
	(0.010)	(0.002)	(0.010)	(0.002)	(0.010)	(0.002)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	728332	728332	725435	725435	728332	728332
Adjusted R^2	0.135	0.133	0.135	0.133	0.135	0.133

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

Table 7: Angel Investor Diversity and Angel Funding

The table shows results from the firm-year level regressions of angel funding on dissimilarity (as in Table 4) interacting with the diversity of local angel investors. In each set of two columns, *Dissimilarity* is measured using the original, vintage, soft-cosine, and dynamic versions as described in Section 3. *High diversity* is an MSA-year level indicator variable that is equal to one if the diversity measure (based on *Education* or *Demographics*) is above the median and zero otherwise. Each diversity measure is computed as the sum of the standardized diversity components by its mean and standard deviation, where each diversity component is calculated as 1-(HHI of a component). *Education* comprises 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 B.

Panel A: Demographic diversity

	Original		Vintage		Soft-cosine	
	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)
Dissimilarity×High diversity	0.027**	0.006*	0.029**	0.006*	0.026	0.005
	(0.012)	(0.003)	(0.013)	(0.003)	(0.017)	(0.004)
Dissimilarity	0.319***	0.069***	0.546***	0.120***	0.632***	0.139***
	(0.082)	(0.020)	(0.161)	(0.039)	(0.171)	(0.041)
High diversity	-0.025**	-0.005*	-0.027**	-0.006*	-0.025	-0.005
	(0.012)	(0.003)	(0.013)	(0.003)	(0.016)	(0.004)
Ln(Age)	0.002*	0.001***	0.002**	0.001***	0.002**	0.001***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Ln(Firms)	-0.054***	-0.013***	-0.054***	-0.013***	-0.052***	-0.013***
	(0.010)	(0.002)	(0.011)	(0.002)	(0.010)	(0.002)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	659320	659320	656423	656423	659320	659320
Adjusted R^2	0.133	0.131	0.134	0.132	0.134	0.132

Panel B: Education diversity

	Original		Vintage		Soft-cosine	
	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding	$\mathbb{1}$(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)
Dissimilarity×High diversity	0.033**	0.007*	0.035**	0.008*	0.040**	0.009**
	(0.016)	(0.004)	(0.017)	(0.004)	(0.018)	(0.004)
Dissimilarity	0.382***	0.081***	0.614***	0.136***	0.704***	0.156***
	(0.084)	(0.020)	(0.166)	(0.040)	(0.177)	(0.042)
High diversity	-0.030*	-0.007*	-0.032**	-0.007*	-0.037**	-0.008*
	(0.016)	(0.004)	(0.016)	(0.004)	(0.017)	(0.004)
Ln(Age)	0.002	0.001***	0.002*	0.001***	0.002*	0.001***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Ln(Firms)	-0.059***	-0.014***	-0.060***	-0.014***	-0.057***	-0.014***
	(0.010)	(0.002)	(0.010)	(0.002)	(0.010)	(0.002)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	611373	611373	608476	608476	611373	611373
Adjusted R^2	0.133	0.132	0.133	0.132	0.133	0.132

Standard errors in parentheses

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

Table 8: 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 diversity as portfolio characteristics. For those measures, we consider all fundings made by each investor during the entire sample period. *Dissimilar* is an indicator variable if the investor's investment portfolio is identified as following 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 Investor* 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 investors' MSA level. For variable definitions and further details of their construction, see Internet Appendix B.

	𝟙(Subsequent Funding)	Exit Rate - IPO	Exit Rate - IPO/Acq.
	(1)	(2)	(3)
Angel Investor	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 Investor	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 9: Angel Investors and Startup Entry Decision

The table shows results from the regressions of firm entry decisions on business dissimilarity and angel investor participation. The sample consists of firm-spatial unit-entry year level observations of 81,718 entering firms over the 2007–2018 sample period. *Entry* is an indicator variable that is equal to one if a given spatial unit area is an actual entering location and zero otherwise. There are 416 MSAs, 149 CSAs, and 1,983 randomly chosen counties in the sample. *Dissimilarity* is one minus the similarity score. The regressions use the elimination of the general solicitation prohibition of Rule 506(c) as an exogenous shock to angel investors' participation. *Treated* is a state-level indicator variable that is equal to one if the number of accredited investors during the pre-shock period is above the median and zero otherwise. *PostGS* is an indicator variable that is equal to one after the general solicitation shock in 2013 and zero otherwise. The sample excludes years before 2011 to mitigate any confounding effects of the 2010 Dodd-Frank amendment of the accredited investor eligibility. All columns control for the local economic variables. 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 B.

	Entry × 100		
	MSA	CSA	County
	(1)	(2)	(3)
Dissimilarity×Treated×PostGS	0.2213**	1.1816***	0.0287*
	(0.108)	(0.353)	(0.015)
Dissimilarity	-0.0985**	-2.5370***	0.0061
	(0.043)	(0.137)	(0.005)
Treated× PostGS	-0.2505**	-1.4290***	-0.0387**
	(0.106)	(0.346)	(0.016)
Ln(Firms)	0.0885***	-1.1914***	-0.0911***
	(0.013)	(0.037)	(0.003)
Ln(Firm Deaths)	0.0275***	1.8155***	0.0244***
	(0.008)	(0.032)	(0.001)
Ln(Investors)	0.3018***	0.4149***	0.0323***
	(0.002)	(0.006)	(0.001)
Ln(GDP)	-0.0701***	-1.2155***	0.0777***
	(0.006)	(0.012)	(0.002)
Net jobs created	-0.1291***	-0.1172***	-0.0136***
	(0.002)	(0.005)	(0.001)
Housing Price Index	0.0033***	0.0256***	0.0013***
	(0.000)	(0.000)	(0.000)
Top college 50	0.2636***	1.3431***	-0.0490***
	(0.007)	(0.014)	(0.004)
Observations	16176848	5486656	11458806
Adjusted R^2	0.022	0.046	0.007
State FE	Y	Y	Y
Firm FE	Y	Y	Y
Entry-year FE	Y	Y	Y

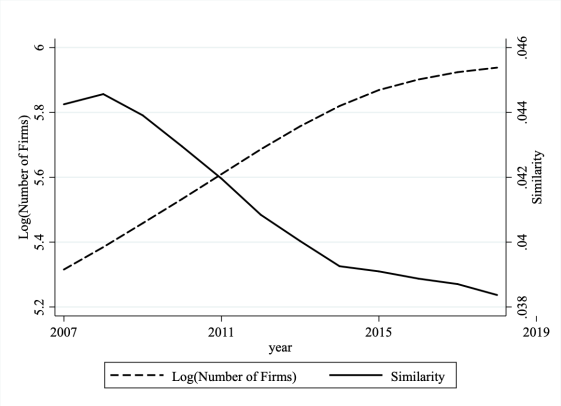
Standard errors are in parentheses.
***p<0.01, **p<0.05, *p<0.1

Internet Appendix to:

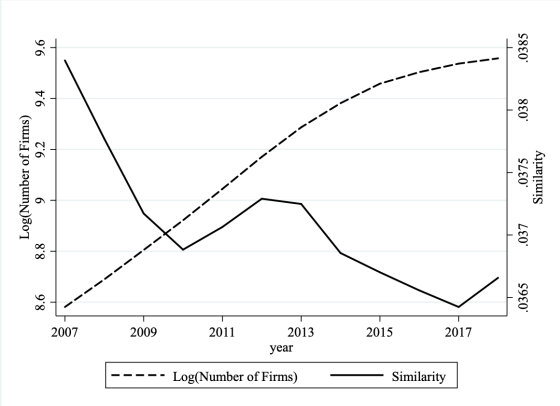
Who Finances Disparate Startups?

Figure IA.1: Similarity Scores over Time

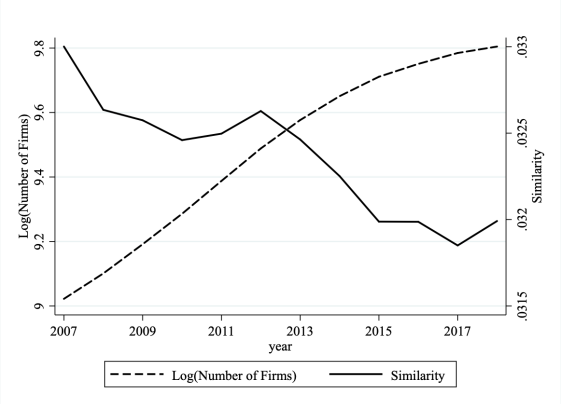
The figures show the time trends of firm similarity scores and the number of firms for all MSAs in Panel (a) and the top 3 active startup activity MSAs, which are San Francisco, New York, Los Angeles, in Panels (b), (c), and (d), respectively. The solid line is our similarity measure. The short dashed line is the total number of firms. For variable definitions and further details of their construction, see Internet Appendix B.



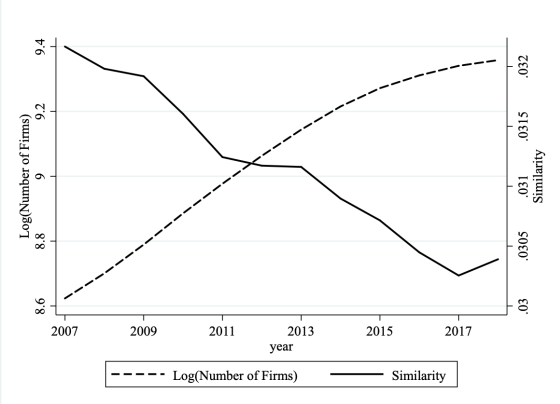
(a) All MSAs



(b) San Francisco MSA



(c) New York MSA



(d) Los Angeles MSA

Table IA.1: Geographic Distribution of Crunchbase Firms

The table shows the location distribution of firms in Crunchbase. The left panel uses the final sample firms between 2007 and 2018 after applying the sample selection described in Section 2.2. The right panel uses all firms 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	117,944			Total	243,237		

Table IA.2: Crunchbase Category Trends

The table presents the 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.

	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 Investor Characteristics

The table presents summary statistics on angel investor characteristics. There are 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 (46% of the entire angel investors)						
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: Angel vs. VC Portfolio Preferences

The table shows results from the regressions of investor portfolio characteristics on investor type. Regressions 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)-(3) and the geographic distance in columns (4)-(6). For those measures, we consider all funded firms during the entire sample period by each investor in Panel A and previously funded firms by each investor before making a given funding in a given year in Panel B. The sample in columns (1) and (4) consists of all investor types, and the sample in columns (2), (3), (5) and (6) consists of only VC and angel investors. *Angel Investor* and *VC Investor* are indicator variables. $Ln(Investments)$ is the log of the total number of investments, $Ln(Investing\ Years)$ is the log of the years since the first investment, and $\$Funding$ is the average funding amount. $Ln(Investments)$ 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 B.

Panel A: Investor-level

	Industry Similarity			Geographic Distance		
	(1)	(2)	(3)	(4)	(5)	(6)
Angel Investor	-0.001	-0.032***	-0.039***	-85.059*	-90.501***	-90.277**
	(0.011)	(0.011)	(0.009)	(43.849)	(34.140)	(36.354)
VC Investor	0.034***			14.769		
	(0.011)			(25.221)		
Ln(Investments)	-0.061***	-0.066***		6.830	7.584	
	(0.005)	(0.006)		(7.655)	(7.737)	
Ln(Investing Years)	-0.020	-0.054***	-0.016	-23.141	-12.844	-55.410
	(0.018)	(0.016)	(0.020)	(56.602)	(65.206)	(70.099)
$Funding	0.000	0.001**	0.001***	4.311***	4.940***	4.922***
	(0.000)	(0.000)	(0.000)	(0.576)	(1.209)	(1.190)
Investor MSA FE	Y	Y	Y	Y	Y	Y
Num. investment	N	N	Y	N	N	Y
Observations	7185	6070	5952	11853	10384	10267
Adjusted R^2	0.094	0.101	0.116	0.068	0.060	0.048

Panel B: Funding-level

	Industry Similarity			Geographic Distance		
	(1)	(2)	(3)	(4)	(5)	(6)
Angel Investor	-0.035***	-0.038***	-0.035***	-140.341***	-72.876***	-69.241***
	(0.007)	(0.005)	(0.004)	(41.823)	(19.804)	(22.389)
VC Investor	0.003			-65.093*		
	(0.006)			(36.859)		
Ln(Investments)	-0.013***	-0.016***		34.327***	44.341***	
	(0.002)	(0.002)		(9.695)	(12.002)	
Ln(Investing Years)	0.020***	0.018***	0.038	123.790***	136.473***	144.884***
	(0.007)	(0.005)	(0.023)	(24.146)	(46.449)	(49.652)
$Funding	0.001***	0.001***	0.001***	5.534***	4.216**	4.512***
	(0.000)	(0.000)	(0.000)	(1.143)	(1.691)	(1.153)
Funding year FE	Y	Y	Y	Y	Y	Y
Investor MSA FE	Y	Y	Y	Y	Y	Y
Funded MSA FE	Y	Y	Y	Y	Y	Y
Num. investment FE	N	N	Y	N	N	Y
Observations	52229	45020	45019	51428	44353	44352
Adjusted R^2	0.226	0.236	0.327	0.088	0.087	0.133

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

Table IA.5: Angel vs. VC Portfolio Preferences – Robustness

The table shows results from the regressions of investor portfolio characteristics on investor type in Table IA.4 by the number of investment bins. Regressions in Panels A and B are at the aggregate investor level, and the sample consists only of VC and angel investors. We consider within investor industry category similarities in Panel A and the geographic distance in Panel B. For those measures, we consider all funded firms during the entire sample period by each investor. *Angel Investor* and *VC Investor* are indicator variables. $Ln(Investments)$ is the log of the total number of investments, $Ln(Investing\ Years)$ is the log of the years since the first investment, and $\$Funding$ is the average funding amount. Standard errors are clustered at the investor MSA level. For variable definitions and further details of their construction, see Internet Appendix B.

Panel A: Industry similarity

	2-11	12-21	22-31	32 or greater
	(1)	(2)	(3)	(4)
Angel Investor	-0.032***	-0.061***	-0.064***	-0.045***
	(0.012)	(0.006)	(0.018)	(0.009)
Ln(Investments)	-0.151***	-0.026	0.027	-0.030***
	(0.013)	(0.020)	(0.082)	(0.002)
Ln(Investing Years)	0.025	-0.038	-0.076***	-0.038***
	(0.038)	(0.024)	(0.021)	(0.012)
$Funding	0.000	0.001*	0.001*	0.001***
	(0.000)	(0.001)	(0.000)	(0.000)
Investor MSA FE	Y	Y	Y	Y
Observations	3769	841	351	1009
Adjusted R^2	0.091	0.122	0.079	0.146

Panel B: Geographic Distance

	2-11	12-21	22-31	32 or greater
	(1)	(2)	(3)	(4)
Angel Investor	-116.359**	-116.635**	-119.698	-77.372**
	(47.935)	(47.779)	(97.113)	(31.411)
Ln(Investments)	2.853	-182.985	-76.872	36.741***
	(18.378)	(113.283)	(192.380)	(10.557)
Ln(Investing Years)	-51.873	116.962	-109.207	144.023**
	(103.225)	(162.887)	(196.828)	(64.097)
$Funding	4.402***	6.003***	2.296	5.563***
	(1.245)	(1.351)	(1.629)	(1.236)
Funding year FE	Y	Y	Y	Y
Investor MSA FE	Y	Y	Y	Y
Funded MSA FE	Y	Y	Y	Y
Observations	4449	843	350	1005
Adjusted R^2	0.066	0.186	0.173	0.239

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

Table IA.6: Propensity Score Matching

The table shows the balancing test of the propensity score matching of treated and control states. Panels A and B each report the balancing test result of the state-level and MSA-level matching, respectively. We match each treated state or MSA to the two nearest matched states or MSAs by the propensity scores estimated using the pre-period (2007-2012) mean of the covariates. *Top school* 50 is a MSA-level cross-sectional variable and cannot be included in the propensity score estimation due to the collinearity with the treatment variable. Columns (1) and (2) are the means of the unmatched sample, columns (3) and (4) are the means of the matched sample, and column (5) presents the p-value of the mean difference ((3)-(4)) of the matched sample. For variable definitions and further details of their construction, see Internet Appendix B.

Panel A: State-level

	Unmatched		Matched		
	Treated	Control	Treated	Control	p-value
	(1)	(2)	(3)	(4)	(5)
Ln(Firms)	4.115	3.688	4.044	4.018	0.924
Ln(Firm Deaths)	6.683	6.406	6.625	6.614	0.953
Ln(Investors)	1.901	1.324	1.824	1.794	0.908
Ln(Population)	-0.448	-0.779	-.506	-.521	0.926
Ln(GDP)	3.409	3.034	3.346	3.308	0.842
Ln(Establishments)	9.515	9.230	9.465	9.436	0.870
Ln(MedHHIncome)	3.873	3.931	3.930	3.937	0.770
HPI growth	-0.041	-0.035	-0.039	-0.043	0.513
Top school 50	0.295	0.092	0.267	0.228	0.548

Panel B: MSA-level

	Unmatched		Matched		
	Treated	Control	Treated	Control	p-value
	(1)	(2)	(3)	(4)	(5)
Ln(Firms)	3.987	1.449	1.668	1.933	0.399
Ln(Firm Deaths)	6.709	5.605	5.649	5.923	0.223
Ln(Investors)	1.622	0.074	0.092	0.064	0.732
Ln(Population)	-0.550	-1.535	-1.578	-1.344	0.219
Ln(GDP)	3.300	2.069	2.078	2.324	0.190
Ln(Establishments)	9.462	8.415	8.417	8.682	0.163
Ln(MedHHIncome)	3.911	3.714	3.734	3.742	0.850
HPI growth	-0.031	-0.018	-0.015	-0.010	0.520

Table IA.7: The validity of the general solicitation shock effects

The table shows results from the regressions that examine the effects of the general solicitation shock on the total angel and VC funding using the MSA-year level observations. We use the elimination of the general solicitation prohibition of Rule 506(c) as an exogenous shock to angel investors' participation. *Treated* is an indicator variable that is equal to one if the number of accredited investors during the pre-shock period is above the medium and zero otherwise. *PostGS* is an indicator variable that is equal to one after the general solicitation shock in 2013 and zero otherwise. GS_{t+n} indicates years from the general solicitation shock in 2013, which is the reference year. The sample excludes years before 2011 to mitigate any confounding effects of the 2010 Dodd-Frank amendment of the accredited investor eligibility. Columns (1) through (4) use the full sample, whereas columns (5) through (8) use the propensity score matched sample, using two nearest neighbors. Control variables are included in the regressions but unreported in the table to conserve space. Standard errors are clustered at the state level. For variable definitions and further details of their construction, see Internet Appendix B.

	Full sample				Matched sample			
	Total angel funding		Total VC funding		Total angel funding		Total VC funding	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated×PostGS	6.387**		-2.667		6.529**		-84.308	
	(2.430)		(24.166)		(2.806)		(91.606)	
Treated×GS_{t-2}		-0.022		25.600		3.445		42.110
		(1.922)		(17.268)		(4.647)		(33.005)
Treated×GS_{t-1}		-0.778		22.310		3.600		49.327
		(1.905)		(13.901)		(5.225)		(32.865)
Treated ×GS_{t+1}		3.817		12.033		5.302		-25.524
		(2.513)		(15.708)		(4.481)		(43.317)
Treated ×GS_{t+2}		6.740**		21.747		9.463**		-6.081
		(2.796)		(14.924)		(4.314)		(30.937)
Treated ×GS_{t+3}		4.280**		-12.452		7.133*		-56.176
		(1.713)		(29.112)		(4.227)		(70.502)
Treated ×GS_{t+4}		5.301**		22.831		9.075**		-75.527
		(2.390)		(25.498)		(4.076)		(101.594)
Treated ×GS_{t+5}		10.114**		20.416		13.270***		-98.308
		(4.218)		(29.294)		(4.568)		(119.099)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1707	1707	1707	1707	2325	2325	2325	2325
Adjusted R^2	0.433	0.431	0.458	0.456	0.589	0.588	0.685	0.687

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

Table IA.8: Angel Investors and Local Startup Similarity (full sample)

The table shows results from reestimation of Table IA.8 regressions that examine the effects of the general solicitation shock on local startup industry diversity using the full-sample. We use the elimination of the general solicitation prohibition of Rule 506(c) as an exogenous shock to angel investors' participation. *Similarity* is the business similarity score. The state-level treatment sample excludes years before 2011 to mitigate any confounding effects of the 2010 Dodd-Frank amendment of the accredited investor eligibility. The state-level *Treated* is an indicator variable that is equal to one if the number of accredited investors during the pre-shock period is above the median and zero otherwise. The MSA-level treatment sample is unaffected by the Dodd-Frank amendment and uses full sample period between 2007 and 2018. The MSA-level *Treated* is an indicator variable that is equal to one if the percentage of households with income greater than $200,000 during the pre-shock period is above the medium and zero otherwise. *PostGS* is an indicator variable that is equal to one after the general solicitation shock in 2013 and zero otherwise. Control variables are lagged by one year. Standard errors are clustered at the state level. For variable definitions and further details of their construction, see Internet Appendix B.

	Similarity × 100					
	State-level treatment			MSA-level treatment		
	Original	Vintage	Soft-cosine	Original	Vintage	Soft-cosine
	(1)	(2)	(3)	(4)	(5)	(6)
Treated×PostGS	-0.312**	-0.344**	-0.318**	-1.136**	-0.892*	-0.813*
	(0.147)	(0.142)	(0.146)	(0.504)	(0.494)	(0.479)
Ln(Firms)	0.125	0.025	0.088	0.360	0.415*	0.421*
	(0.165)	(0.228)	(0.256)	(0.258)	(0.235)	(0.240)
Ln(Firm Deaths)	0.581	0.626	0.637	-0.157	0.103	0.079
	(0.874)	(0.768)	(0.796)	(0.327)	(0.309)	(0.312)
Ln(Investors)	0.152	0.208	0.255*	-0.030	-0.080	-0.061
	(0.148)	(0.138)	(0.150)	(0.140)	(0.133)	(0.135)
Ln(Population)	-0.620	-0.759	-0.674	3.096	1.945	1.642
	(0.846)	(0.789)	(0.821)	(2.388)	(2.360)	(2.437)
Ln(GDP)	0.737	0.598	0.579	0.137	-0.274	-0.120
	(0.580)	(0.489)	(0.502)	(0.711)	(0.752)	(0.741)
Ln(Establishments)	-1.340	-1.054	-1.277	-2.451	-0.896	-0.878
	(1.342)	(1.217)	(1.242)	(2.144)	(2.088)	(2.105)
Ln(MedHHIncome)	1.393	1.495	1.610	1.120	1.481	1.519
	(1.092)	(1.116)	(1.135)	(0.841)	(0.920)	(0.935)
HPI growth	0.329	0.149	0.215	0.331	0.444	0.448
	(0.907)	(0.902)	(0.937)	(0.579)	(0.500)	(0.497)
Top school 50	0.446*	0.532**	0.558**			
	(0.251)	(0.245)	(0.258)			
State FE	Y	Y	Y	N	N	N
MSA FE	N	N	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	1707	1707	1707	2221	2221	2221
Adjusted R^2	0.180	0.195	0.213	0.780	0.769	0.787

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

Table IA.9: Treatment effects over time

The table shows results from the regressions that examine the effects of the general solicitation shock on local startup industry diversity using the MSA-year level observations. We use the elimination of the general solicitation prohibition of Rule 506(c) as an exogenous shock to angel investors' participation. *Treated* is an indicator variable that is equal to one if the number of accredited investors during the pre-shock period is above the medium and zero otherwise. GS_{t+n} indicates years from the general solicitation shock in 2013, which is the reference year. The sample excludes years before 2011 to mitigate any confounding effects of the 2010 Dodd-Frank amendment of the accredited investor eligibility. Columns (1) through (3) use the full sample, whereas columns (4) through (6) use the propensity score matched sample, using two nearest neighbors. Control variables are included in the regressions but unreported in the table to conserve space. Standard errors are clustered at the state level. For variable definitions and further details of their construction, see Internet Appendix B.

	Similarity × 100					
	Full sample			Matched sample		
	Original	Vintage	Soft-cosine	Original	Vintage	Soft-cosine
	(1)	(2)	(3)	(4)	(5)	(6)
Treated×GS_{t-2}	-0.0128	0.0329	0.0405	0.2037	0.2162*	0.2193*
	(0.188)	(0.176)	(0.185)	(0.132)	(0.121)	(0.125)
Treated×GS_{t-1}	0.0594	0.0443	0.0400	0.0849	0.0847	0.0762
	(0.117)	(0.111)	(0.117)	(0.087)	(0.078)	(0.081)
Treated ×GS_{t+1}	-0.1049	-0.0287	-0.0280	-0.0585	0.0161	0.0051
	(0.068)	(0.109)	(0.110)	(0.083)	(0.115)	(0.116)
Treated ×GS_{t+2}	-0.3628**	-0.4216**	-0.3999**	-0.5102**	-0.5480*	-0.5494**
	(0.164)	(0.199)	(0.192)	(0.248)	(0.279)	(0.263)
Treated ×GS_{t+3}	-0.2407	-0.3242*	-0.2989*	-0.3501**	-0.3829**	-0.3857**
	(0.164)	(0.174)	(0.175)	(0.159)	(0.188)	(0.167)
Treated ×GS_{t+4}	-0.3446**	-0.4177***	-0.3774***	-0.4450**	-0.4920**	-0.4864**
	(0.144)	(0.135)	(0.131)	(0.179)	(0.199)	(0.184)
Treated ×GS_{t+5}	-0.4087**	-0.3779***	-0.3332**	-0.4062**	-0.3756***	-0.3704***
	(0.165)	(0.130)	(0.139)	(0.163)	(0.131)	(0.135)
Controls	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	1707	1707	1707	2270	2270	2270
Adjusted R^2	0.178	0.194	0.211	0.502	0.464	0.514

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

Table IA.10: Business Dissimilarity and VC Funding

The table shows the results from the firm-year level regressions of VC funding on dissimilarity, analogous to the main regressions in Table 4. Columns (1) and (2) use the full sample by replacing no-funding observations with zeros. Columns (3) and (4) use a subsample of funding observations only. $\mathbb{1}(VC)$ 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. For variable definitions and further details of their construction, see Internet Appendix B.

	$\mathbb{1}$(VC)	\$Funding	$\mathbb{1}$(VC)	\$Funding
	(1)	(2)	(3)	(4)
Dissimilarity	-0.139	-2.849**	-2.154**	-203.228***
	(0.111)	(1.319)	(0.998)	(75.066)
Ln(Age)	0.026***	0.226***	0.139***	3.183***
	(0.004)	(0.051)	(0.018)	(0.442)
Ln(Firms)	-0.054***	0.015	-0.186**	40.385***
	(0.014)	(0.078)	(0.089)	(4.726)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	728332	728332	21151	21151
Adjusted R^2	0.237	0.243	0.437	0.491

Table IA.11: Business Dissimilarity and Angel Funding – Robustness

The table shows results from the firm-year level regressions of angel funding on dissimilarity for robustness tests using the funding subsample. Panel A supplements data with Form-D filings. Panel B drops funding observations reported more than 2 years after the funding announcement date to mitigate concerns with the selection of successful funding. In each set of two columns, *Dissimilarity* is measured using the original, vintage, soft-cosine, and dynamic versions as described in Section 3. $\mathbb{1}(Angel)$ 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 B.

Panel A: Supplementing with Form-D filings

	Original		Vintage		Soft-cosine	
	𝟙(Angel)	$Funding	𝟙(Angel)	$Funding	𝟙(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)
Dissimilarity	5.466***	22.416*	6.744***	32.382**	6.731***	34.055**
	(1.235)	(11.777)	(1.221)	(14.674)	(1.217)	(14.055)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	30800	30800	30763	30763	30830	30830
Adjusted R^2	0.376	0.213	0.376	0.213	0.376	0.220

Panel B: Dropping late funding report observations

	Original		Vintage		Soft-cosine	
	𝟙(Angel)	$Funding	𝟙(Angel)	$Funding	𝟙(Angel)	$Funding
	(1)	(2)	(3)	(4)	(5)	(6)
Dissimilarity	3.775***	6.867***	3.738***	6.311**	3.780***	6.366**
	(0.940)	(2.500)	(0.919)	(2.609)	(0.843)	(2.563)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	19336	19336	19318	19318	19336	19336
Adjusted R^2	0.361	0.296	0.361	0.296	0.361	0.296

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

Table IA.12: Business Dissimilarity and Entrepreneurship Quality

The table presents the MSA-year-level regressions of the Entrepreneurship Quality Index (EQI) on MSA-level business dissimilarity. The dependent variable is the MSA-level EQI measure from Andrews, Fazio, Guzman, Liu, and Stern (2022), which is available up to 2013. Standard errors are clustered at the MSA level. For variable definitions and further details of their construction, see Internet Appendix B.

	EQI×100		
	(1)	(2)	(3)
MSA Dissimilarity	-0.007	-0.007	-0.003
	(0.010)	(0.010)	(0.013)
Ln(Firms)		-0.012	-0.066*
		(0.013)	(0.038)
Ln(Investors)		0.002	-0.001
		(0.002)	(0.003)
Ln(Personal Income)		0.036***	0.005
		(0.013)	(0.010)
Regional Entrepreneurship Cohort Potential Index			0.004***
			(0.000)
Regional Ecosystem Acceleration Index			-0.000***
			(0.000)
Observations	3430	3271	1983
Adjusted R^2	0.843	0.844	0.786
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.13: The List of Emerging Industries

The table presents the list of 38 emerging industries identified by ChatGPT among 742 Crunchbase categories, specifically related to our sample period from 2007 to 2018.

3D Printing	Augmented Reality	Cloud Computing
3D Technology	Autonomous Vehicles	Cyber Security
Ad Retargeting	Blockchain	Data Mining
Ad Targeting	Clean Energy	Digital Entertainment
Advanced Materials	EdTech (Education Technology)	Energy Efficiency
Aerospace	Environmental Consulting	FinTech (Financial Technology)
AgTech (Agriculture Technology)	GreenTech	Health Diagnostics
Alternative Medicine	Internet of Things (IoT)	Machine Learning
Artificial Intelligence	Medical Device	Mobile Apps
Augmented Reality	Nanotechnology	Renewable Energy
Autonomous Vehicles	Robotics	Semantic Web
Blockchain	Smart Building	Smart Cities
Clean Energy	Smart Home	Social Media Management
Cloud Computing	Virtual Reality	Wearables

Internet Appendix A 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
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 B 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.
Dissimilarity	One minus the similarity score.
Industry Similarity	A within-investor industry similarity of portfolio firms.
Geographic Distance	A distance in the mile between an investor and a funded firm.
Angel (VC)	An indicator variable equal to one if funding is raised from angel investors (VCs) and zero otherwise.
$ Total Funding	The total funding amount received in a year (in millions).
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.
Angel Investor	An indicator variable equal to one if a given investor's Crunchbase investor type includes angel and zero otherwise.
VC Investor	An indicator variable equal to one if given investor's Crunchbase investor type includes venture capital and zero otherwise.
Treated	An indicator variable, which is equal to one if the state-level number of accredited investors is above median and zero otherwise. The post-Dodd Frank amended definition of accredited investor is used.
PostGS	An indicator variable equal to one after the elimination in 2013 and zero otherwise.
Ln(Firms)	The natural logarithm of the number of firms in Crunchbase.
Ln(Firm Deaths)	The natural logarithm of the number of closed firms from Business Dynamics Statistics.
Ln(Investors)	The natural logarithm of the number of investors in Crunchbase.
Ln(Population)	The natural logarithm of population obtained from the Bureau of Economic Analysis.
Ln(GDP)	The natural logarithm of the real gross domestic product obtained from the Bureau of Economic Analysis.
Ln(Establishments)	The natural logarithm of the number of establishments from Business Dynamics Statistics.
HPI growth	The annual growth of quarterly all-transactions index (indexed to 1995). The data is obtained from the Federal Housing Finance Agency.
Net jobs create	An indicator variable, which is equal to one if there was a positive number of net jobs created and zero otherwise. It captures the change in employment due to opening/closing establishments or expanding/contracting establishments. The data is obtained from the Business Dynamics Statistics from the public Census data.
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.
Dissimilar	An indicator variable equal to one if an investor's portfolio consists of firms with below median average similarity and zero otherwise.
Entrepreneurship Quality Index (EQI)	MSA-year level average entrepreneurial quality obtained from Andrews et al. (2022). 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. (2022). 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 obtained from Andrews et al. (2022). The data is available between 1988-2013.
Top college	An indicator variable, which equals one if a U.S. top 50 college in 2008 is located in the MSA and zero otherwise. The data is obtained from U.S. News.
High Diversity	A MSA-year level angel diversity indicator variable equal to one if the diversity measure (*Education* or *Demographics*) is above the median and zero otherwise. Each diversity measure is computed as the sum of normalized diversity components 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 follows: 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 ($(1/3)^2 + (2/3)^2$).
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.