Intermediating Ideas*

Vrinda Mittal S. Lakshmi Naaraayanan Charles Parry

July 31, 2026

[Preliminary, Please Do Not Circulate]

Abstract

Venture capital may not be fully fungible across entrepreneurial ideas: investors' prior experience can shape which start-ups they are willing or able to finance. Using business descriptions for a near-universe of U.S. venture-backed firms, we construct a high-dimensional idea space and locate each venture capitalist based on its prior portfolio. We establish three facts. First, venture capitalists invest disproportionately near areas in which they have previously invested. Second, investors with broader portfolios reach farther in idea space, but less than proportionately to their greater breadth, indicating that broader experience expands but does not eliminate the limits of prior specialization. Third, investors' locations in idea space affect which ideas receive funding. Guided by a model of investor choice, we use the Volcker Rule as an exogenous contraction in capital supply among exposed venture capitalists. Following the shock, a one-standard-deviation increase in an idea region's accessibility to affected investors reduces the number of VC-backed firms. The effect shows no differential pre-trends and is robust to controlling for geographic exposure. Our findings highlight that the distribution of investor experience— not only the aggregate supply of venture capital—shapes financing across entrepreneurial ideas.

*Mittal: Department of Finance, Kenan-Flagler Business School, University of North Carolina at Chapel Hill (Email: Vrinda_Mittal@kenan-flagler.unc.edu, Web: https://www.vrindamittal.com/); Naaraayanan: London Business School and CEPR (Email: lnaaraayanan@london.edu, Web: https://www.lakshmin.com); and Parry: Imperial College London (Email: c.parry@imperial.ac.uk, Web: https://sites.google.com/view/charles-parry). Naaraayanan thanks London Business School's Research and Materials Development Grant and Wheeler Institute for Business and Development for supporting this research. We thank Dennis Zhao for his outstanding research assistance. All errors are our own.

1 Introduction

Entrepreneurial firms are central to creative destruction and the implementation of new technologies (Schumpeter, 1942; Aghion and Howitt, 1990). Because new ideas generate economic value only when they are developed and commercialized, access to finance is an important determinant of long-run innovation and growth (King and Levine, 1993; Aghion et al., 2018). Entrepreneurs often lack the resources to pursue their ideas independently and therefore rely on venture capital. Although venture capital finances only a small fraction of new firms, VC-backed companies account for a disproportionate share of successful and innovative businesses (Kerr et al., 2014; Bloom et al., 2020; Gornall and Strebulaev, 2021). Therefore, its allocation may shape not only how much entrepreneurial activity is financed, but also which ideas receive the opportunity to be developed and commercialized.

Yet venture capital may not be fully fungible across entrepreneurial ideas. Venture capitalists do more than supply funding: they screen, monitor, and support their portfolio companies, activities that may draw on experience with particular technologies, markets, and business models (Kaplan and Strömberg, 2001; Lerner and Nanda, 2020; Bonelli, 2025; Ewens and Malenko, 2026). This gives rise to two views of how funding is allocated. Under the first, capital follows commercial opportunity: investors accumulate experience wherever promising start-ups emerge, so proximity between their new and prior investments simply reflects the clustering of high-quality opportunities. Under the second, accumulated experience itself shapes where funding flows. If evaluating and supporting unfamiliar ventures is costly, investors may favor start-ups similar to their prior portfolios even when more distant opportunities offer comparable commercial potential. In that case, the allocation of capital depends not only on the distribution of opportunities, but also on the distribution of investor experience across the venture capital market.

In this paper, we distinguish between these views by placing start-ups and investors in a common idea space. Using business descriptions for the near-universe of U.S. venture-backed firms in PitchBook, we represent each company as a point in a high-dimensional space, where nearby firms pursue similar products, technologies, markets, or business models. We locate each venture capitalist using its prior portfolio and measure the breadth of that portfolio within the same space. An investor's prior portfolio summarizes where it has accumulated experience, relationships, and investment capacity, while also reflecting factors such as investment mandates, partner preferences, networks, and deal flow. This continuous representation allows us to study similarity across firm activities without imposing discrete

industry boundaries or conflating proximity in ideas with geographic proximity. We then use this measure to ask whether investors' prior location in the idea space shapes the allocation of venture funding.

We establish three novel facts on the growth of ideas from the lens of VC financing. First, venture capitalists invest disproportionately near areas of idea space in which they have previously invested. Holding fixed the set of start-ups funded in a given year, actual investments are systematically closer to investors' prior portfolios than placebo investments formed by randomly reallocating start-ups across investors, including within sector. Second, investors with broader portfolios reach farther from their prior investments, but less than proportionately to their greater breadth. Broader prior experience therefore expands investors' reach in idea space without eliminating the importance of where they have invested before. Third, investors' locations in idea space affect which ideas receive funding.

These patterns are consistent with investor experience shaping funding, but they are not by themselves sufficient to establish this interpretation. Investors may have accumulated portfolios near areas where high-quality opportunities persist, perhaps because knowledge spillovers produce clusters of related start-ups. The same clustering would generate local investment even if capital moved frictionlessly across idea space. Because the quality and location of unfunded opportunities are unobserved, separating investor-side frictions from the "spatial" distribution of opportunities is a key identification challenge.

To address this challenge, we build a stylized model of VC investment in which investors face a stochastic discrete-choice problem over startups, trading off firm quality against the cost of investing outside their domain of expertise, with that cost governed by the investor's portfolio breadth. The model rationalizes both stylized facts as equilibrium outcomes of investor expertise, and delivers testable predictions for how the allocation of funding across the idea space should respond to an exogenous shock to capital supply, independent of any change in the underlying quality of ideas.

We test this prediction using the Volcker Rule as a regulatory shock to the capital available to exposed venture capitalists (Chen and Ewens, 2025). We partition idea space into equal-sized patches using a deterministic k-d tree constructed solely from pre-shock firm locations. We then combine investors' exposure to the rule with their pre-shock accessibility to each patch, as determined by idea-space distance and portfolio breadth. Exposure is therefore greater where more strongly affected investors were better positioned to provide funding before the rule. The resulting shift-share design compares changes in funding across patches with different predetermined exposure to the same capital-supply shock. The primary identifying assumption is that, absent the rule, funding in more- and less-exposed patches

would have evolved similarly, conditional on the set of fixed effects. The empirical specification also includes Patch fixed effects that absorb persistent differences across regions of idea space and year fixed effects that absorb aggregate changes in venture funding.

We also find that funding falls disproportionately in regions of idea space that were more exposed to affected investors before the rule. A one-standard-deviation increase in exposure is associated with an 11 percent decline in the number of firms receiving VC funding after 2014. The estimate is essentially unchanged after controlling for an analogous measure of geographic exposure, indicating that the cross-patch pattern is not explained by physical proximity between investors and start-ups. The result is consistent with capital held by investors in different parts of idea space being imperfectly substitutable over the relevant horizon. Together, the evidence suggests that who holds venture capital, and where those investors have previously invested, shapes which entrepreneurial ideas receive financing.

Although these findings do not by themselves imply that ideas farther from existing investors are underfunded in a welfare sense— we do not observe the quality of unfunded opportunities or the social returns to alternative allocations of capital— they do establish an important prior step. Namely, the allocation of venture funding reflects the location and breadth of intermediaries in idea space, not only the underlying distribution of commercial opportunities. Our results therefore speak to the incidence of changes in venture-capital supply rather than to the efficiency of the resulting allocation. In particular, aggregate venture-capital supply may be an incomplete measure of the financing available within a given region of idea space: a contraction affecting a particular set of investors can change where funding flows, not only how much capital is supplied. The composition and prior positioning of intermediaries may therefore matter alongside the total quantity of entrepreneurial capital for both the financing and the direction of innovation.

Related Literature. This paper contributes to three strand of the literature. First, it relates to work on the generation, diffusion, implementation, and economic consequences of new ideas (Bloom et al. (2020), Kalyani (2022), Garfinkel et al. (2023), Paine (2025), Ma (2025), Horton and Sariri (2026), Carvalho and Freund (2026), Crouzet et al. (2026), Fieler and Lee (2026), Ganguli et al. (2026)). We contribute by constructing a continuous idea space from firm-level business descriptions and using it to study the allocation of entrepreneurial finance. Whereas this literature primarily asks where ideas originate, how they diffuse, or which firms implement them, we ask whether the prior positioning of intermediaries affects which ideas receive capital.

Second, our paper contributes to the literature on venture-capital selection, matching, and involve-

ment. Venture capitalists screen and select firms (Sørensen, 2007; Gompers et al., 2020), actively monitor their portfolio companies (Hellmann and Puri, 2002; Bernstein et al., 2016), and add value through networks, portfolio spillovers, and managerial intervention (Hochberg et al., 2007; Ewens and Marx, 2018; González-Uribe, 2020; Ewens and Malenko, 2026). Prior work also shows that entrepreneur characteristics, including education, gender, and race, affect access to and selection into entrepreneurial finance (Lyonnet and Stern, 2026; Ewens and Townsend, 2018; Cook et al., 2026; Fairlie et al., 2022). Related evidence on early-stage investors documents substantial investor-specific heterogeneity and links stronger performance to startup- and industry-specific knowledge, founder connections, and active governance (Karlsen et al., 2024). Relative to existing work, we provide new micro-level evidence on equilibrium matching between start-ups and venture capitalists, which we argue is influenced in part by the prior experience and expertise of investors. Our findings lend support to the view that venture capitalists are active investors, whose knowledge and expertise shape deal selection.[1]

Third, the paper speaks to the broader literature on entrepreneurial finance and the implications of changes in the supply of venture-capital. Research on new firms emphasizes the importance of external finance and shows that young firms' responses to investment opportunities depend on access to capital (Robb and Robinson, 2014; Adelino et al., 2017). Prior work further shows that shocks to venture-capital supply affect start-up activity (Kortum and Lerner, 2000; Samila and Sorenson, 2011; Chen and Ewens, 2025), studies how venture capital contributes to wealth inequality (Gocmen et al., 2025), and examines cross-country differences in venture activity (Parry, 2025). Related research also considers the capacity of venture capital to fund novel or technologically demanding innovations (Nanda and Rhodes-Kropf, 2013; Lerner and Nanda, 2020; Bonelli, 2025). Our paper shows that the existing stock of venture capital expertise is an important determinant of the allocation of funding to start-ups pursuing different ideas and technologies: VCs tend to invest in familiar areas, and this local bias mediates how shocks to venture capital funding are spread across different technologies. In this sense, our work relates to papers that study the capacity of the venture capital market to fund truly novel innovations (see, Nanda and Rhodes-Kropf, 2013; Bloom et al., 2020; Lerner and Nanda, 2020).

[1]Our paper also contributes to the literature on the trade-off between exploitation and exploration. Prior work shows that exploratory innovation is shaped by incentives and financing arrangements that tolerate early failure, reward long-run success, and lower the cost or financing risk of experimentation (Manso, 2011; Azoulay et al., 2009; Tian and Wang, 2014; Kerr et al., 2014; Nanda and Rhodes-Kropf, 2013; Ewens et al., 2018); related evidence shows that accumulated industry experience affects how venture capitalists respond to changing investment opportunities (Gompers et al., 2008). We complement this literature by showing that the exploration margin also depends on where investors are positioned in idea space: investors with broader portfolios reach farther from their prior investments, but less than proportionately, so the distribution of investor experience shapes which unfamiliar ideas receive funding.

The remainder of the paper proceeds as follows. Section 2 describes the data and our text-based measurement of idea space and novelty. Section 3 documents the stylized facts on VC investment and firm novelty. Section 4 develops the model and derives testable implications. Section 5 describes the empirical implementation of the model's predictions and presents results. Section 6 concludes.

2 Data and Summary Statistics

The data on firms and deals is sourced from Pitchbook. Pitchbook provides detailed information on the entrepreneurial companies financed and the underlying deals. Company information includes a company identifier, name, primary industry classification, and a business description, that serves as the input for all subsequent text analysis. The deal information includes transaction records detailing the type of deal, deal size, and firm identifiers, which allows us to link the deal characteristics to the firms.

The sample is filtered to retain entrepreneurial companies. To do so, a deal (and correspondingly the firm) is kept if any of the deal-type fields are among the following types: Accelerator/Incubator, Angel (Individual), Restart-Angel, Seed Round, Early-Stage VC, Restart – Early VC, Later Stage VC, Restart – Later VC, or PE Growth/Expansion. The filtered deals are then merged with the company characteristics, retaining only the matched firm-deal observations. Further, the sample is restricted to firms headquartered in the U.S. The time period studied is 1980-2025, with a focus on the post 2000s. Additionally, deal sizes are aggregated within a firm across deal types to produce a single aggregate deal size per firm. Hence, we end up with the U.S. firms which are entrepreneurial in nature, financed by venture capital and angel backed investors, broadly defined.

Table 1 reports the distribution of firms across the seven primary industry sectors in the final sample. Information Technology is the largest sector, accounting for 37.84 percent of the 136,082 firm-deal observations and spanning 38 industry codes across 6 industry groups. In total, the sample covers 219 industry codes across 41 industry groups within 9 industry sectors.

We embed each firm's business description into a 768-dimensional vector using the all-mpnet-base-v2 sentence-transformer model of Reimers and Gurevych (2019). This places every company as a point in a common high-dimensional idea space, where firms with semantically similar descriptions lie close together. We measure the proximity of two firms by the cosine similarity of their embeddings. Next, for each firm we compute a similarity score — the maximum cosine similarity between its embedding and those of same-sector firms founded in the prior ten years. The similarity score captures how closely a

Table 1: Primary Industry Sector Distribution with Industry Groups and Codes

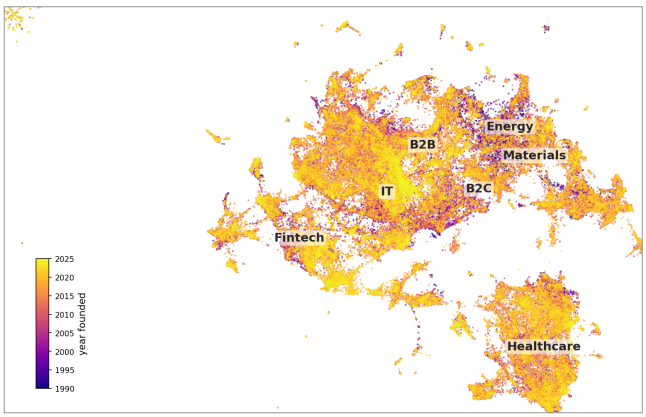
Sectors	Freq.	Percent	Groups	Codes
Business Products and Services (B2B)	24,096	17.71	4	31
Consumer Products and Services (B2C)	27,140	19.94	9	46
Energy	3,491	2.57	5	19
Financial Services	4,865	3.58	4	24
Healthcare	22,600	16.61	5	26
Information Technology	51,491	37.84	6	38
Materials and Resources	2,399	1.76	8	35
Total	136,082	100.00	41	219

firm's idea resembles the most similar preceding firm in its sector. Higher values indicate lower novelty, while lower values indicate a firm whose idea is more distinct from what came before.

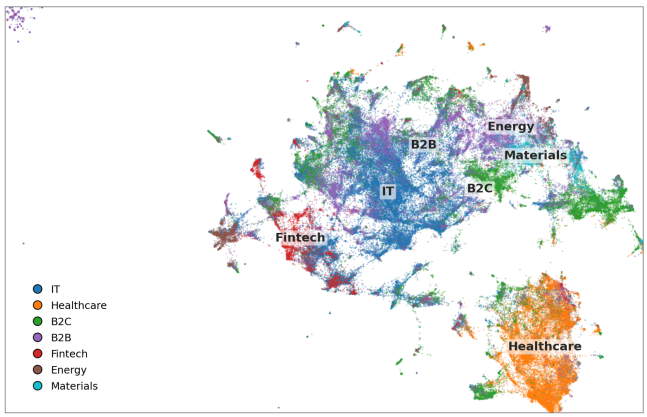
Figure 1 illustrates the idea space. Since the embeddings are 768-dimensional, for display purposes, we compress them to two dimensions using UMAP (Uniform Manifold Approximation and Projection). All measures use the full-dimensional embeddings. Panel A colors each firm by founding year, and Panel B colors firms by primary sector. Firms cluster into recognizable sectors, with healthcare apart and Information Technology, Fintech, and the broad B2B and B2C sectors overlapping at the center. Comparing Panel A and Panel B, we observe that the most recent foundings concentrate in Information Technology and Fintech.

Figure 1: The Idea Space of The Entrepreneurial Sample

(A) By Founding Year



(B) By Sector



Notes: The figure plots the idea space of the sample by transforming the 768-dimensional space embeddings to two dimensions using UMAP (Uniform Manifold Approximation and Projection). Panel A shows different colors based on founding year, and Panel B is by industry.

Table 2 reports the distribution of this score across the seven primary industry sectors. Across the 128,911 firms in the sample, the score averages 0.778, with a median of 0.793 and a standard deviation of 0.099, and it ranges from 0.048 to 1. Financial Services and Information Technology have the least novel

firms on average, with means of 0.793 and 0.789, respectively. Materials and Resources stands out as the most novel sector, with a substantially lower mean of 0.706, and Healthcare sits modestly below the rest at 0.767.

Table 2: Descriptive Statistics of The Firm-level Similarity Score By Sector

Sectors	Obs.	Mean	Median	SD	Min	Max
Business Products and Services (B2B)	26,552	0.780	0.788	0.108	0.055	1.000
Consumer Products and Services (B2C)	32,728	0.777	0.793	0.096	0.141	1.000
Energy	3,758	0.776	0.795	0.106	0.048	1.000
Financial Services	5,608	0.793	0.810	0.094	0.150	1.000
Healthcare	23,879	0.767	0.780	0.095	0.103	1.000
Information Technology	33,825	0.789	0.805	0.093	0.085	1.000
Materials and Resources	2,561	0.706	0.718	0.110	0.132	1.000
Total	128,911	0.778	0.793	0.099	0.048	1.000

Figure 2 plots the 10-year rolling average of pairwise similarity scores by sector over time. The shaded area represents the inter quartile range. The panel restricts the incumbents founded within the prior ten years. The similarity score within each sector increases over time, which suggests that more similar ideas are founded relative to startups instituted in the past ten years. In Appendix Figure A3, we decompose this increase in similarity scores by restricting the firms each year to a random subset equal to the number of firms in 2000. This removes the possibility of a mechanical increase due to more number of firms instituted in the recent years. Despite the slight decrease in the upward trend in similarity scores, the substantial increase over time still persists.

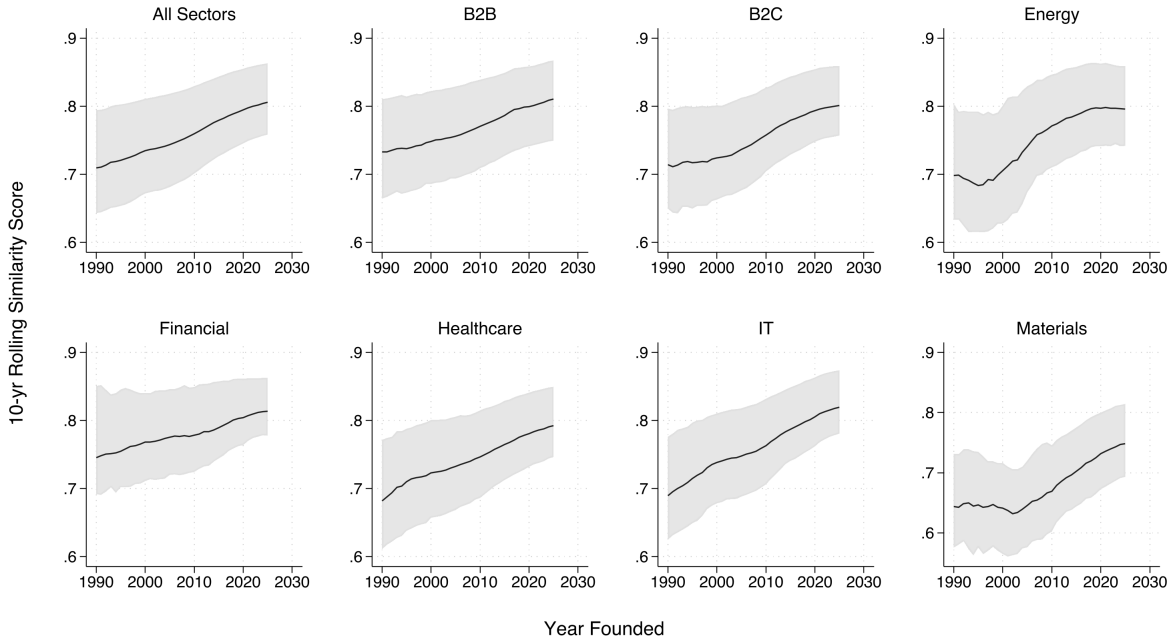
3 Stylized Facts

This section documents two stylized facts about VC investment behavior. To do so, we utilize the business descriptions in Pitchbook, which we transform into embeddings and represent by a unit-length vector $\mathbf{x}_i \in \mathbb{R}^n$, locating it in an n-dimensional "ideas space", $\mathcal{X}$. The construction of these embeddings is described in section 2.

Each firm i is then represented by a vector $\mathbf{x}_i$ and an initial financing date, t_i. Similarly, each VC j may be represented by a portfolio of prior investments. We use $e_{ij} = 1$ to indicate that investor j was the lead investor in the first-financing round for firm i, with $e_{ij} = 0$ otherwise. VC j is then summarized by a portfolio

$$M_{jt}^{\tau} = \{i \mid e_{ij} = 1, t_i \in [t - \tau, t)\},$$

Figure 2: Idea Novelty over Entrant Founding Year Across Industries



Notes: The figure shows similarity scores over entrant founding year across industries. Each point shows the 10-year rolling average similarity score by sector. Shaded area represents the interquartile range (P25-P75) of pairwise similarity scores.

which collects the firms that received investment from VC j in the τ years prior to year t. In practice, we restrict attention to the portfolio of VC j over the prior five years, M_{jt}^5, to focus on a more up-to-date sample of portfolio firms. We summarize the location of the VC's portfolio using its mean portfolio vector

$$\mathbf{m}_{jt} = \frac{1}{|M_{jt}^5|} \sum_{i \in M_{jt}^5} \mathbf{x}_i, \tag{1}$$

and, because firms are represented by a vector $\mathbf{x}_i$ of unit length, the mean portfolio vector lies weakly within the unit sphere ,$||\mathbf{m}_{jt}|| \leq 1$. The direction of the mean portfolio vector

$$\hat{\mathbf{m}}_{jt} = \frac{\mathbf{m}_{jt}}{||\mathbf{m}_{jt}||}$$

identifies the central location of VC j's recent investment experience in idea space, while its magnitude $||\mathbf{m}_{jt}||$ captures the concentration of that experience. Specifically, we define portfolio breadth

$$b_{jt} = 1 - ||\mathbf{m}_{jt}||, \tag{2}$$

which is equivalent to the average cosine distance of the portfolio firms from the portfolios average

direction.[2] Intuitively, when VC j invests in a narrow set of firms, its mean portfolio vector will lie near to the unit sphere, whereas when VC j's investments differ substantially, the mean portfolio vector will lie closer to the origin.

With these firm and VC descriptions, we then define two measures of the distance between a start-up i and VC j. First, the average distance

$$\bar{d}_{ijt} = 1 - \mathbf{x}_i \cdot \hat{\mathbf{m}}_{jt} \tag{3}$$

is the cosine distance between start-up i and the average direction of VC j's prior investments.

Second, as a robustness, we define the nearest-neighbor distance

$$d_{ijt}^{NN} = 1 - \max_{\iota \in M_{jt}^5} \mathbf{x}_i \cdot \mathbf{x}_\iota, \tag{4}$$

which is the cosine distance from startup i to the most similar firm in the portfolio of VC j. Relative to the average distance, $\bar{d}_{ijt}$, the nearest-neighbor distance is robust to VCs having multiple narrow areas of investment that are separated in the ideas space. In such cases, a startup may be far from the portfolio's average location while remaining close to at least one of the VC's prior investments.

Using these simple descriptions of start-ups and VCs, we now proceed to document several stylized facts about the micro-level investment behavior of VCs.

Fact 1 – VC investments exhibit local bias in ideas space. Figure 3 compares the distribution of distances between VCs j and the investments i to a placebo distribution, in which each VC investment is reallocated to a randomly chosen start-up i that received investment in that year. Panel A does this for the mean distance, $\bar{d}_{ijt}$, and Panel B does this for the nearest-neighbor distance, d_{ijt}^{NN}. In each case, the placebo start-up i is sampled from both the full set of start-ups receiving investment in that year, and a restricted set of start-ups that also belong to the same sector.

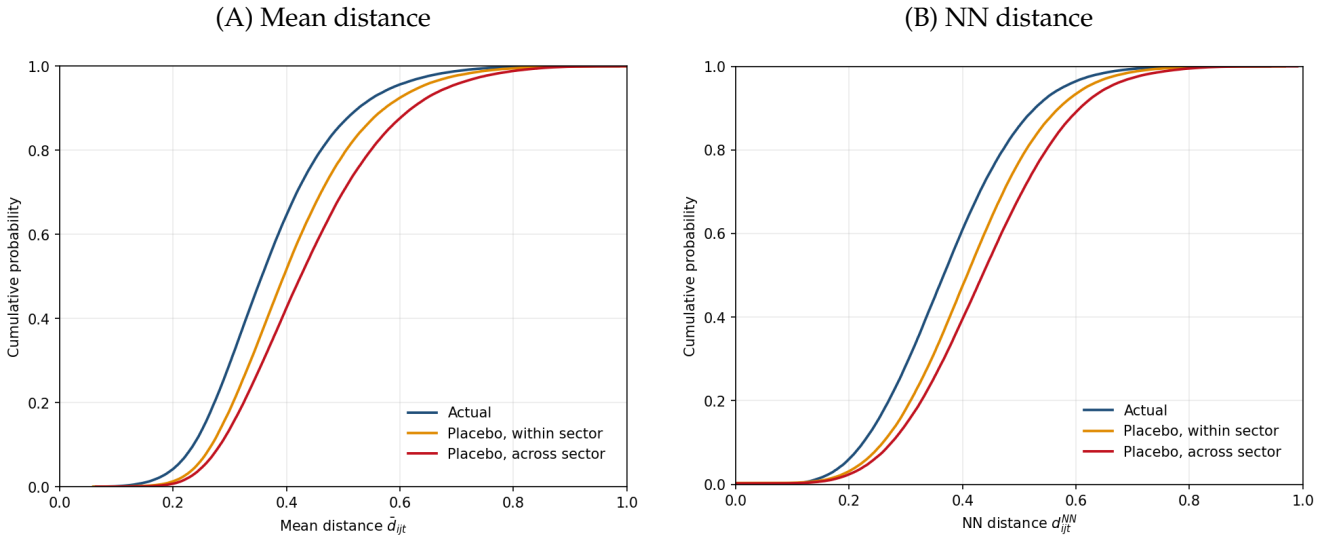
In each case, the actual distribution of distances is smaller – in a first-order stochastic dominance sense – than the placebo distribution, which demonstrates that VCs favor new investments in similar areas to their previous investments. To the extent that past investments proxy for domain expertise, the findings suggest that VCs invest in areas close to their existing expertise. In practice, this accumulated

[2] The average cosine distance

$$\frac{1}{|M_{jt}^5|} \sum_{i \in M_{jt}^5} \left(1 - \frac{\mathbf{x}_i \cdot \mathbf{m}_{jt}}{||\mathbf{x}_i|| \, ||\mathbf{m}_{jt}||}\right) = 1 - \frac{1}{|M_{jt}^5|} \sum_{i \in M_{jt}^5} \left(\mathbf{x}_i \cdot \frac{\mathbf{m}_{jt}}{||\mathbf{m}_{jt}||}\right) = 1 - \left(\frac{1}{|M_{jt}^5|} \sum_{i \in M_{jt}^5} \mathbf{x}_i\right) \frac{\mathbf{m}_{jt}}{||\mathbf{m}_{jt}||} = 1 - \frac{\mathbf{m}_{jt} \cdot \mathbf{m}_{jt}}{||\mathbf{m}_{jt}||} = 1 - ||\mathbf{m}_{jt}||.$$

experience from investing in a certain area may reflect improved screening and monitoring efficiency, or a better understanding of the distribution of underlying payoffs associated with investments in a particular domain. As robustness, Appendix Figure A2 shows the results from a similar exercise in which, rather than constructing the placebo by reassigning the VC to a new firm, the placebo is constructed by reallocating the firm to a new VC. This exercise delivers the same qualitative findings.

Figure 3: VC investments exhibit local bias.



Notes: CDFs of the idea-space distance from a VC to its new investment in the new-company sample. Curves show the actual distance and two placebos that reassign the investment to a random start-up funded that year, within or across sector. Panel A uses the mean distance to the VC's prior five-year portfolio and Panel B the nearest-neighbor distance.

Fact 2 – Greater portfolio breadth is associated with more distant investments, but the association is less than proportionate. Figure 4 shows how the distance between a VC and its new investment varies with the measured breadth of the VC's prior portfolio. Specialist investors — those with narrower prior portfolios — tend to make new investments that are closer, in absolute terms, to their prior investments than do VCs with broader existing portfolios.

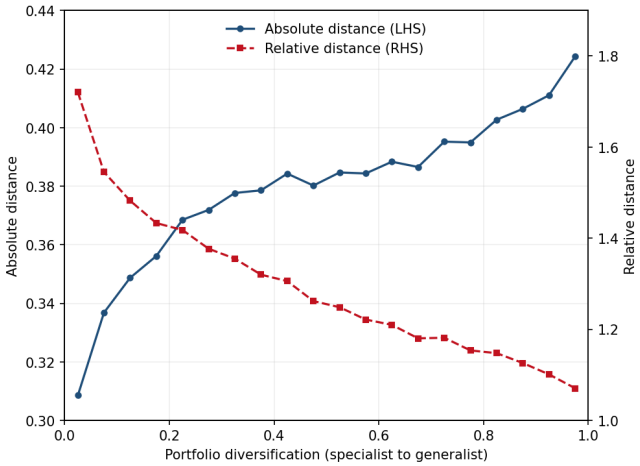
However, the figure also shows that the positive association between portfolio breadth and the distance of new investments is less than proportionate. To illustrate this, we normalize each distance measure by the breadth of the VC's prior portfolio, b_{jt}, and refer to the resulting measure as relative distance. In contrast to absolute distance, relative distance declines with portfolio breadth.

These findings speak to how investment expertise scales across the idea space. VCs with broader prior portfolios — the "generalist" investors — make investments that are farther from their existing portfolios in absolute terms, but closer relative to the span of their prior investments. This suggests that a broader base of existing expertise increases a VC's ability to invest in startups that are new-to-the-VC,

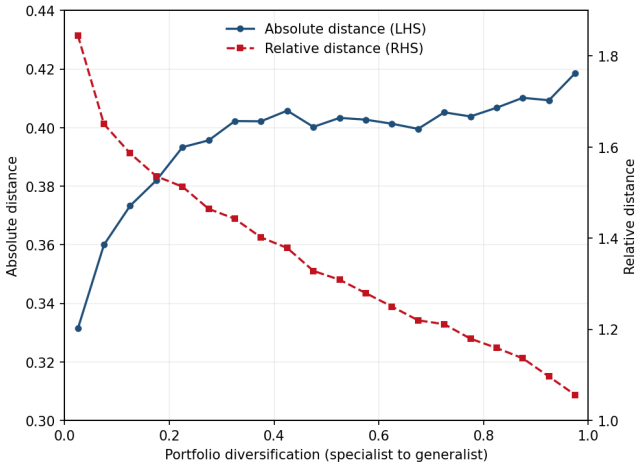
but that this ability increases less than proportionately with the breadth of its existing knowledge.

Figure 4: Portfolio diversification and investment distance: absolute vs relative.

(A) Mean distance



(B) NN distance



Notes: VC-year investments in the new-company sample, binned into 20 groups by portfolio diversification from specialist to generalist. Absolute distance measures the distance from the new investment to the VC's prior five-year portfolio, and relative distance divides it by the portfolio's breadth b_{jt}, equal to $\bar{d}_{ijt}/b_{jt}$ in Panel A and d^{NN}_{ijt}/b_{jt} in Panel B.

4 Model

The stylized facts are consistent with the existing stock of VC expertise playing an important role in the allocation of funding across technologies: funding allocation may not simply reflect the distribution of commercial opportunities, but may also be shaped by the market structure of the VC industry. In this environment, "novel" ideas may be underfunded, relative to their commercial potential, because they are less likely to lie within the domain expertise of existing VCs.

However, this interpretation is difficult to establish without a measure of the underlying distribution of commercial opportunities. The observed technological proximity of VC investments may simply reflect the spatial distribution of startup opportunities or technological spillovers across related areas. In that case, the patterns documented above would be driven by the supply of ideas rather than by the supply of venture capital.

This section develops a simple model of VC investment motivated by the stylized facts in Section 3. In the model, VC expertise play a crucial role in shaping the allocation of funding, and the model delivers testable predictions that result from this interpretation of the micro-level facts about VC investment behavior. Section 5 subsequently finds evidence for these predictions empirically, which serves to support our interpretation of the facts.

4.1 Model setup

The model is static. There is a continuous ideas space, $\mathcal{X} \subseteq \mathbb{R}^k$ endowed with a distance metric, $d(.,.)$. Within this ideas space, there is a finite set of VCs, indexed by $j \in \mathcal{J} = \{1, 2, ..., J\}$, that seek to make investments, and a finite set of start-ups, indexed by $i \in \mathcal{I} = \{1, 2, ..., I\}$, that require investment.

Each start-up i is endowed with a idea location $x_i \in \mathcal{X}$ and a quality, $q_i \in \mathbb{R}$. In turn, VCs enter the period with an existing portfolio, which we summarize by a location, $m_j \in \mathcal{X}$ and a breadth, $b_j > 0$, which is intended to capture the scope the VC's existing investments across the ideas space.

The investment decision Each VC j is limited to investing in at most one start-up i. Before seeking a start-up, each VC j receives a capital supply shock $a_j \in \{0, 1\}$, which determines whether it has capital available to invest. If $a_j = 1$, investor j is active and is otherwise inactive. We impose $\Pr(a_j = 1) = \pi_j$, and assume that capital supply shocks are independent across investors and independent of the match-specific shocks introduced below.

Conditional on being active, $a_j = 1$, investor j observes each start-up i and is aware of the associated payoff

$$u_{ij} = q_i - \frac{\kappa}{b_j^{\eta}} d(m_j, x_i) + \varepsilon_{ij}, \qquad \kappa, \eta > 0, \tag{5}$$

from investing in start-up i. The first term captures start-up quality, which is independent of the investor's identity. The second term captures the cost associated with "exploration" for the VC, which may be motivated through a lack of information, higher monitoring and screening costs, or other factors, and the parameter η determines how portfolio breadth affects the cost of exploration. Finally, the third term, ε_{ij}, is a match-specific component, which is independently and identically distributed Type-I extreme value across investors and start-ups.

Each investor seeks to invest in the start-up that maximizes u_{ij}

$$i_j^* = \arg\max_{i \in \mathcal{I}} u_{ij}$$

and the distributional assumption on the match-specific component, ε_{ij}, ensures that each VC has a unique preferred start-up. Furthermore, the probability that start-up i is strictly preferred by investor j is given by

$$p_{ij} = \frac{\exp\{q_i - \frac{\kappa}{b_j^\eta} d(m_j, x_i)\}}{\sum_{\iota \in \mathcal{I}} \exp\{q_\iota - \frac{\kappa}{b_j^\eta} d(m_j, x_\iota)\}}, \tag{6}$$

which is a function of its quality, its position in the ideas space vis-à-vis the investor, and the strength of competing investment opportunities, through the denominator.

Equation (6) gives the probability that start-up i is optimal for investor j, but i may also be optimal for other investors k. Therefore, to map from VC preferences to investment outcomes, we need to impose a restriction to prevent two VCs from choosing the same start-up.

Assumption 1. *The opportunity set, $\mathcal{I}$, is sufficiently large and diffuse that, for fixed J, $\lim_{I\to\infty} \bar{p}_I = 0$, where $\bar{p} = \max_{i,j} p_{ij}$ is the maximum choice probability among all start-up–investor pairs.*

Under Assumption 1, the probability that any two VCs choose the same start-up i vanishes as the opportunity set grows.[3] Equation (6) then gives the true investment probabilities.

Aggregation To move from individual investment behavior to market-level outcomes, we now aggregate individual VC choices across the ideas space. Let $\mathcal{A} \subseteq \mathcal{X}$ be any region, or "patch", of the ideas space and let

$$\mathcal{I}(\mathcal{A}) = \{i \in \mathcal{I} \mid x_i \in \mathcal{A}\}$$

be the set of start-ups in that patch. The probability that an active investor j selects a start-up in $\mathcal{A}$ is given by

$$P_j(\mathcal{A}) = \sum_{i \in \mathcal{I}(\mathcal{A})} p_{ij},$$

which simply sums across the probabilities of each individual start-up in patch $\mathcal{A}$. The following proposition aggregates these investment probabilities to determine the number of investments in patch $\mathcal{A}$.

Proposition 1. *Under Assumption 1, each VC chooses a unique start-up i, so the expected number of investments in a patch $\mathcal{A}$ is given by*

$$\mathbb{E}[N(\mathcal{A})] = \sum_{j \in \mathcal{J}} \pi_j P_j(\mathcal{A})$$

[3] To see this, note that the probability that some start-up $i \in \mathcal{I}$ is optimal for both investors j and k is $\Pr(i_j^* = i_k^*) = \sum_{i \in \mathcal{I}} p_{ij} p_{ik}$. Then, since $p_{ij} \leq \bar{p}$ (by definition), $\Pr(i_j^* = i_k^*) = \sum_{i \in \mathcal{I}} p_{ij} p_{ik} \leq \bar{p} \sum_{i \in \mathcal{I}} p_{ik} = \bar{p}$, where the final equality follows from $\sum_{i \in \mathcal{I}} p_{ik} = 1$.

and the probability that patch $\mathcal{A}$ experiences at least one investment is given by

$$\Pr(N(\mathcal{A}) \geq 1) = 1 - \Pr(N(\mathcal{A}) = 0) = 1 - \prod_{j \in \mathcal{J}} \left[1 - \pi_j P_j(\mathcal{A})\right].$$

Proposition 1 is the main aggregation result. The number of investments in a given patch of the ideas space depends on several factors. First, it depends naturally on the quality of the ideas, q_i for $i \in \mathcal{I}(A)$, in that patch: all else equal, higher-quality ideas attract more funding. Second, it depends on the position of the patch relative to investors. Holding fixed the quality of the ideas in the patch, funding is higher when the patch lies close to investors' existing portfolios.

Third, funding in a patch depends on the alternatives available to the same investors. An attractive neighboring patch reduces funding in $\mathcal{A}$ when the investors who might otherwise fund $\mathcal{A}$ also regard that neighboring patch as a plausible investment opportunity. This force is stronger when the relevant investors are broad generalists, because their wider reach makes them more able to substitute across different patches of the ideas space. Specialists, by contrast, allocate their investment probability more locally, so competition from distant patches is weaker.

4.2 The model and the micro facts

In this subsection, we show that the stylized facts arise as natural implications of the model under mild conditions.

Fact 1 – VC investments exhibit local bias in ideas space Towards demonstrating local investment bias in the ideas space, fix a VC j and order start-ups by their distance d_{ij} from j in $\mathcal{X}$

$$d_{(1)j} \leq d_{(2)j} \leq \cdots \leq d_{(I)j},$$

so that (i) refers to the start-up with the i-th smallest distance from j. Then, from (6), the probability that VC j's actual investment lies no farther away than the n-th closest start-up, $d_{(n)j}$, is

$$F_j^A(d_{(n)j}) = \sum_{i=1}^{n} p_{(i)j} = \frac{\sum_{i=1}^{n} \exp\{q_{(i)} - \frac{\kappa}{b_j^\eta} d_{(i)j})\}}{\sum_{\iota \in \mathcal{I}} \exp\{q_{(\iota)} - \frac{\kappa}{b_j^\eta} d_{(\iota)j}\}} \tag{7}$$

Now, suppose instead that the VC were randomly assigned to invest in some start-up i, where assignment is uniform. Then, the probability that VC j's placebo investment lies no farther away than the

n-th closest start-up, $d_{(n)j}$, is

$$F_j^P(d_{(n)j}) = \sum_{i=1}^{n} \frac{1}{I} = \frac{n}{I} \tag{8}$$

Comparing equations (7) and (8) demonstrates that there are two forces shaping the ranking of $F_j^A(d_{(n)j})$ and $F_j^P(d_{(n)j})$. On the one hand, for given quality, start-ups located near to investor j are more likely investments for j, which pushes towards $F_j^A(d_{(n)j}) > F_j^P(d_{(n)j})$. On the other hand, investment probabilities also respond to commercial opportunities, and this may overturn the ranking between $F_j^A(d_{(n)j})$ and $F_j^P(d_{(n)j})$ at some distance $d_{(i)j}$. For instance, if VC j were located in a part of the ideas space containing many low quality start-ups, they may be inclined to invest far from their existing expertise to access better opportunities. Together, this suggests that the relationship between the quality of commercial opportunities, q_i, and distance from the VC, d_{ij}, is important. The following lemma provides a sufficient condition such that the actual distribution of investment distances is first-order stochastically smaller than the placebo distribution.

Lemma 1. *Suppose that the quality of start-up (i) satisfies $q_{(i)} = \bar{q} + \alpha d_{(i)j}$. If $\alpha < \frac{\kappa}{b_j^\eta}$, then the actual investment-distance distribution is first-order stochastically smaller than the uniform-placebo distance distribution:*

$$F_j^A(d) \geq F_j^P(d) \qquad \textit{for all } d.$$

Outside the static model, this condition is arguably not especially restrictive. VC location is measured using prior portfolio companies, so that if the spatial distribution of high-quality opportunities is persistent, VCs' inherited locations will tend to be closer to future high-quality opportunities. Furthermore, Lemma 1 refers to the investment distribution for a single VC j, whereas Figure 3 is equivalent to observing one draw from each investor-specific actual distribution, F_j^A, and one draw from each corresponding placebo distribution, F_j^A. Stochastic dominance in the pooled distribution is a weaker requirement than stochastic dominance for every VC j individually.

Fact 2 – Greater portfolio breadth is associated with more distant investments, but the association is less than proportionate Towards analyzing the relationship between portfolio breadth and investment distance, consider two VCs, a specialist s and a generalist g where $b_s < b_g$, that face that same joint distribution of start-up qualities and distances. Then, after ordering start-ups by distance, we have

$$d_{(i)s} = d_{(i)g} = d_{(i)} \qquad \text{and} \qquad q_{(i)s} = q_{(i)g} = q_{(i)}.$$

The actual distribution of distances for each investor $j \in \{s, g\}$ is given by equation (7). The following lemma establishes that the distribution of investment distances for the specialist investor, $j = s$, is first-order stochastically smaller than the distribution for the generalist, $j = g$.

Lemma 2. *Suppose that VCs s and g face the same joint configuration of start-up qualities and distances, but $b_s < b_g$. Then*

$$F_s^A(d) \geq F_g^A(d) \qquad \text{for all } d,$$

so that the specialist's actual investment-distance distribution is first-order stochastically smaller than the generalist's.

Lemma 2 establishes that investment breadth is associated with more distance investments by considering the full distribution of investment distances. To analyse relative distance, it is simpler to consider only how the expected relative distance responds to portfolio breadth. To this end, first note the implication of Lemma 2, that expected absolute distance is increasing in portfolio breadth.

Corollary 1. *The expected (absolute) distance from investor j to its investment is $\mathbb{E}_j^A[d_{ij}] = \sum_{i \in \mathcal{I}} p_{ij} d_{ij}$. It's first derivative with respect to the portfolio breadth of investor j is is given by*

$$\frac{\partial \mathbb{E}_j^A[d_{ij}]}{\partial b_j} = \eta \frac{\kappa}{b_j^{\eta+1}} Var_j^A(d_{ij}) > 0,$$

so that expected absolute distance is increasing in portfolio breadth.

Given Corollary 1, it is simple to see that the derivative of expected relative distance, $E_j^A[d_{ij}/b]$, with respect to portfolio breadth is

$$\frac{\partial E_j^A[d_{ij}/b_j]}{\partial b_j} = \frac{1}{b^2}\left[\eta \frac{\kappa}{b^\eta} Var_j^A(d_{ij}) - \mathbb{E}_j^A[d_{ij}]\right], \tag{9}$$

whose sign depends on the joint distribution of commercial opportunities and distances from investor j. The following result then provides a simple sufficient condition for expected relative distance to decline in breadth.

Lemma 3. *There exists an $\bar{\eta}$ such that, if $\eta \in (0, \bar{\eta})$, then $\frac{\partial E_j^A[d_{ij}/b_j]}{\partial b_j} < 0$ for all $j \in \mathcal{J}$.*

Intuitively, η controls how the attractiveness of an investment at distance d from VC j changes in response to the breadth of investor j: when η is small, the positive effect of breadth on the attractiveness of an investment at distance d is more muted. Thus, although portfolio breadth always increases

the propensity of a VC to invest further away in ideas space (Lemma 2), this relationship is less than proportionate when the benefits of scaling portfolio breadth, η, are sufficiently small.

Bringing these insights together, the model can account for both the local bias in investment behavior and the investment behavior patterns of specialist and generalist investors.

4.3 Testable implications of the model

In this section, we derive testable implications of the model. To this point, we have presented a model in which the stock of VC expertise, and its allocation across the ideas space, contributes significantly to the equilibrium allocation of funding across start-ups. We have also shown that the model is consistent with micro-level evidence on VC investment behavior.

However, there are several other plausible explanations for the patterns we observe. Most notably, the local investment behavior of VCs could simply reflect persistence in the location of high-quality ideas, which may arise due to spillovers that are themselves local in the ideas space.

With this in mind, we develop testable implications of the model, specifically regarding how the distribution of funding should respond to a capital supply shock. To the extent that empirical effects of these shocks are mediated through the channels encoded in the model, they serve to support the view that the stock of VC expertise and its allocation contribute to the equilibrium allocation of funding across start-ups.

A capital supply shock Suppose the probability that investor j is active is given by

$$\pi_j = \bar{\pi}_j - \delta B_j, \qquad \delta > 0, \tag{10}$$

where δ is an adverse shock to funding supply and $B_j \geq 0$ is investor j's exposure to the shock. The effect of the shock on the number of investments in a given patch is summarized by the following lemma.

Lemma 4. *The effect of a capital supply shock on expected investment in patch $\mathcal{A}$ is given by*

$$\frac{\partial \mathbb{E}[N(\mathcal{A})]}{\partial \delta} = -\sum_{j \in \mathcal{J}} B_j P_j(\mathcal{A}).$$

where $P_j(\mathcal{A}) = \sum_{i \in \mathcal{I}(\mathcal{A})} p_{ij}$ is the probability that investor j would invest in patch $\mathcal{A}$ if active.

Intuitively, a certain patch will be adversely affected from a capital supply shock if the shock is

particularly severe among investors that were likely to invest in that patch. Therefore, the effect is most pronounced when there are a large number of investors located near to the patch that are specialists – and therefore very likely to direct their investments toward the patch — or when the patch is isolated in the ideas space, implying that nearby VCs have few alternative patches to invest in.

5 Testing the model's implications

Section 5.1 describes the regression specification, Section 5.2 describes the implementation, and Section 5.3 presents estimates of the testable implications.

5.1 Regression Specification

To test the model's testable implications, we estimate a shift-share specification:

$$N_{r,t} = \alpha_r + \lambda_t + \beta \left(Z_r \times \text{Post}_t\right) + \nu_{r,t}, \tag{11}$$

where $N_{r,t}$ is the number of U.S. companies that received US-VC-backed funding in patch r and year t (2001 to 2025), Z_r is the patch exposure standardized to mean zero and unit variance. To construct Z_r, we use the state-level bank exposure to the Volcker rule, $BankExpo_s$, capturing exogenous regulatory shock to financing constraints on venture capital (VC) firms, made publicly available by Chen and Ewens (2025). We construct a VC-patch level accessibility weight $\omega_{j,r}$, which is determined by the idea-space distance between the VC's pre-shock portfolio and the centroid of the idea patch. Detailed construction of idea space in patches in described in Section 5.2. VC-patch accessibility is $\omega_{j,r} = \exp(-\kappa \frac{d_{j,r}}{\theta_j^\eta})$ for each VC j in patch r, where $d_{j,r} = 1 - c_j \cdot \text{centroid}_r$ is the cosine idea-space distance between VC j's pre-shock portfolio centroid and the patch centroid**, θ_j is VC j's continuous portfolio breadth, κ is the distance-decay coefficient, and η governs how breadth scales that decay. Using this, $Z_r = \sum_j \omega_{j,r} BankExp_{s(j)}$, where all the VCs in a state share the same shock and $\omega_{j,r}$ varies across patches.

Post_t takes the value 1 for $\{t \geq 2014\}$. We include patch fixed effects α_r and year fixed effects λ_t, so Z_r and Post_t are each absorbed and only the interaction is identified. We cluster standard errors by patch. Because $N_{r,t}$ is a count with many zeros, we use a Poisson (PPML) as the main estimator, $\mathbb{E}[N_{r,t}] = \exp(\alpha_r + \lambda_t + \beta\, Z_r \times \text{Post}_t)$.

5.2 Partitioning the idea space into patches

We partition the idea space into discrete patches, using a k-d tree that recursively splits the cloud of company embeddings along coordinate axes. The tree is built on the pre-shock firms, those founded on or before 2013, and the same cuts are then applied to every company, so firms that enter after the shock sort into boundaries that are already fixed. Each cut is placed by the mechanical rules in Appendix C.1, so the partition involves no discretionary choices and is exogenous to the shock.

The partition has two features we exploit. First, it is deterministic and tuning-free, with no cluster centers, target number of clusters, or random seed, so the same data always yields the same patches. Its only parameter is the tree depth, which we fix at 10, and Table A1 reports how the patches behave across depths. Second, because the tree is built solely from the idea embeddings of pre-shock firms and uses no funding, investor, or outcome data, the patch boundaries are predetermined with respect to the shock. We drop the 2.4% of companies with uninformative placeholder descriptions, such as "operating in stealth mode," which would otherwise form a single degenerate patch.

Table 3 Panel A reports the patches and the variables used in regression analyses. With the depth parameter as 10, the tree gives 1,024 patches. By construction each patch holds an equal number of firms — an average of 41.39 when considering pre-shock firms and an average of 129.59 firms once later entrants are added. Following the accessibility weights defined in Section 5.1, each patch's raw exposure aggregates the state-level Volcker shock across VCs by their idea-space accessibility to the patch, $e_r = \sum_j \omega_{j,r}, BankExpo_{s(j)}$. We standardize this across patches, $Z_r = \left(e_r - \bar{e}\right)/\text{sd}(e)$, and the average raw patch exposure e_r is 226.20. Firms and their investors both cluster in space, so the first concern is that the patches only proxy for geography. Panel B provides a first evidence to show that they do not. Among VC-backed firms, two firms share a U.S. state with probability 16.6% when they are in the same patch, barely above the unconditional probability 15.1% of being in the same state. Section 5.3 confirms this formally by adding a physical-distance version of the exposure variable to Equation 11.

5.3 Estimation Results for Testable Implication #1

Table 4 reports the estimates. We rely on PPML, as about 20% of patch-years have zero VC-backed funding. In Column (1), the main estimator, $Z_r \times$ Post has a coefficient of -0.121. A one standard deviation increase in patch exposure corresponds to an 11% decrease in VC-backed firms funded after 2014.[4]

[4]The dependent variable $N_{r,t}$ averages 3.05 VC-backed firms per patch-year. The 11% comes from $100 \times (e^{-0.121} - 1) = 11.4\%$, or about 0.35 firms per patch-year at the mean.

Table 3: Summary statistics

Panel A. Idea-space patches

Variable	Mean	SD	Median	Max	N	$N_{>0}$
Companies per patch, pre-shock	41.394	0.489	41.000	42.000	1,024	1,024
Companies per patch, all	129.585	54.546	115.000	569.000	1,024	1,024
VC-backed firms funded $N_{r,t}$	3.054	3.655	2.000	61.000	25,600	20,402
Idea exposure Z_r	226.199	23.780	230.867	267.182	1,024	1,024
Geo exposure Z_r^{geo}	142.839	10.465	142.149	174.052	1,024	1,024
Shares exposure Z_r^{sh}	0.308	0.005	0.307	0.323	1,024	1,024

Panel B. Probability two firms share a US state

Sample	Firms	$P(\text{same state} \mid \text{same patch})$	$P(\text{same state})$
VC-backed	43,957	0.166	0.151

Notes: Panel A describes the idea-space patches and the regression variables. Companies per patch is reported for the pre-shock firms used to build the partition, which are split into equal-size patches by construction, and for all companies including later entrants. $N_{r,t}$ is VC-backed firms funded in patch r and year t, $N_{>0}$ counts nonzero patch-years, and the three exposures are patch-level and enter the regressions standardized to mean zero and unit variance. Panel B gives, for VC-backed firms, the probability that two firms share a US state, unconditionally and conditional on sharing an idea-space patch.

Figure 5 shows the event study plot five years pre and post the shock. The figure confirms no parallel pre-trends before the 2014, and shows sustained decreases in funding 2014 onwards.

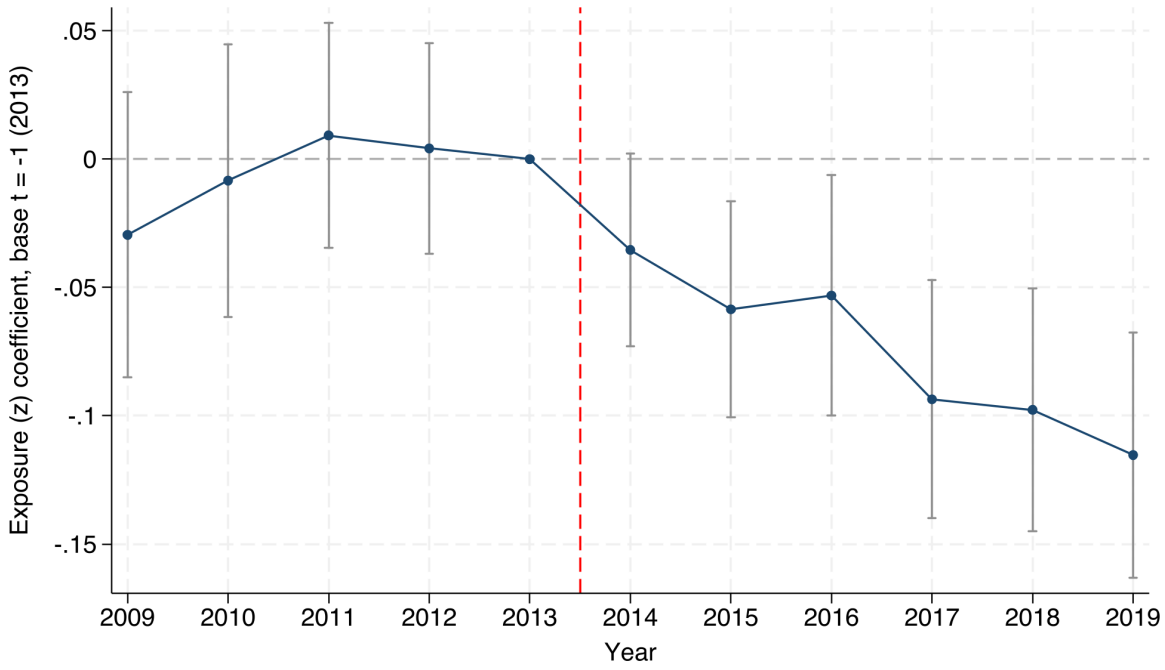
As mentioned above, a natural concern is that the idea-space exposure captures geographic proximity between VCs and firms rather than proximity in ideas. To address this, we construct a physical-distance analogue of the exposure and add it as a control. Specifically, for each VC j and patch r, we compute the great-circle (haversine)[5] distance between the VC's headquarters and each pre-shock firm in the patch, using geocoded ZIP codes, and average these into a single VC-to-patch distance $d_{j,r}^{\text{geo}}$. We then convert this distance into an accessibility $\omega_{j,r}^{\text{geo}} = \exp(-d_{j,r}^{\text{geo}}/h)$, where the bandwidth h is the median VC-to-patch distance, so that closer VCs are more accessible. The geographic exposure is $Z_r^{\text{geo}} = \sum_j B_j \, \omega_{j,r}^{\text{geo}}$, which takes the same shift-share form as the idea-space exposure using the state shocks B_j and U.S. VCs. We add this exposure to the baseline specification,

$$N_{r,t} = \alpha_r + \lambda_t + \beta \, (Z_r \times \text{Post}_t) + \delta \, (Z_r^{\text{geo}} \times \text{Post}_t) + \nu_{r,t},$$

so that β captures the idea-space effect holding physical proximity fixed. Table 4 Column (2) shows that the idea-space coefficient is unchanged when the geographic exposure is added, remaining at -0.121,

[5]The haversine formula uses two points' latitude and longitude to compute the distance between them along the Earth's surface.

Figure 5: VC Funding in Idea Patches Pre and Post Volcker Rule



Notes: The figure shows the β estimates in Equation 11 pre and post the 2014. Standard errors are clustered at the idea patch level. The bars show 90% confidence intervals.

and the geographic exposure itself is small and insignificant at -0.010. Therefore, the idea-space effect does not reflect geography.

Table 4: Idea-space effect and VC funding

	VC Funding	
	(1)	(2)
Idea $Z_r \times$ Post	-0.121^{***}	-0.121^{***}
	(0.027)	(0.027)
Geo $Z_r^{\text{geo}} \times$ Post		-0.010
		(0.027)
Patch FE	Yes	Yes
Year FE	Yes	Yes
Observations	25,575	25,575

Notes: Coefficient estimates use the PPML estimation procedure. Post takes the value one for years 2014 and onwards. Idea Z_r is the idea patch exposure of firms based on firms pre-shock. Geo Z_r^{geo} is the geographic patch exposure of firms based on firms pre-shock. Standard errors are clustered by patch, and displayed in parentheses. Significance levels: *** $p < 0.01$; ** $p < 0.05$, * $p < 0.10$.

6 Conclusion

This paper studies how the expertise accumulated by venture capitalists shapes the allocation of financing across entrepreneurial ideas. Using business descriptions for a near-universe of U.S. venture-backed companies, we construct a high-dimensional idea space for firms and for investors according to their prior portfolios. We document two stylized facts. First, venture capitalists invest disproportionately close to their existing portfolios in idea space. Second, investors with broader portfolios venture farther from their prior investments in absolute terms, but this increase is less than proportional to portfolio breadth. These findings suggest that the ability to finance unfamiliar ideas expands with investor experience, but does not scale frictionlessly.

We formalize this interpretation in a model in which investors trade off startup quality against the costs of screening, monitoring, and supporting firms outside their domains of expertise. The model shows how the distribution of investor expertise can influence funding independently of the underlying distribution of entrepreneurial opportunities. It also predicts that a contraction in capital supply should have its largest effects in the regions of idea space most accessible to the affected investors. We test this prediction using the Volcker Rule as a shock to the capital available to exposed venture capitalists, and estimate a substantial decline in the number of VC-backed firms funded after 2014 in the patch with a high idea-space exposure prior to the shock. The estimates are unchanged after controlling for an analogous measure of geographic exposure.

Taken together, the results show that venture capital is not a fungible input allocated uniformly across technologies. Its effective supply depends on where investor expertise is situated in idea space. Consequently, shocks to financial intermediaries can have sharply uneven effects across technologies, even when the underlying supply and quality of entrepreneurial ideas do not change. Generalist investors partly bridge gaps between domains, but their reach does not eliminate the importance of accumulated, domain-specific knowledge. The market structure of venture capital, that is, the distribution of expertise across investors as well as the aggregate quantity of capital, therefore helps determine which ideas obtain financing.

References

Adelino, Manuel, Song Ma, and David Robinson, "Firm age, investment opportunities, and job creation," *Journal of Finance*, 2017, *72* (3), 999–1038.

Aghion, Levine, Peter Howitt, and Ross Levine, "Financial development and innovation-led growth," in "Handbook of finance and development," Edward Elgar Publishing, 2018, pp. 3–30.

Aghion, Philippe and Peter Howitt, "A model of growth through creative destruction," 1990.

Azoulay, Pierre, Joshua S Graff Zivin, and Gustavo Manso, "Incentives and creativity: evidence from the academic life sciences," Technical Report, National Bureau of Economic Research 2009.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend, "The Impact of Venture Capital Monitoring," *Journal of Finance*, 2016, *71* (4), 1591–1622.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," *American Economic Review*, April 2020, *110* (4), 1104–1144.

Bonelli, Maxime, "Data-driven investors," *Review of Financial Studies*, 2025, p. hhaf078.

Carvalho, Vasco M. and Lukas B. Freund, "Firms as Foragers," Working Paper, University of Cambridge & CEPR and Boston College July 2026.

Chen, Jun and Michael Ewens, "Venture capital and startup agglomeration," *Journal of Finance*, 2025, *80* (4), 2153–2198.

Cook, Lisa D, Matt Marx, and Emmanuel Yimfor, "Funding Black High-Growth Startups," *Journal of Finance*, 2026, *81* (3), 1619–1660.

Crouzet, Nicolas, Zhiguo He, Victor Lyonnet, and Yueran Ma, "Why Don't Old Firms Do New Things?," Technical Report March 2026. Working Paper.

Ewens, Michael and Matt Marx, "Founder replacement and startup performance," *Review of Financial Studies*, 2018, *31* (4), 1532–1565.

_ **and Nadya Malenko**, "Board dynamics over the startup life cycle," *Journal of Finance*, 2026.

_ **and Richard Townsend**, "Are early stage investors biased against women?," *Journal of Financial Economics (JFE), Forthcoming*, 2018.

_ **, Ramana Nanda, and Matthew Rhodes-Kropf**, "Cost of experimentation and the evolution of venture capital," *Journal of Financial Economics*, 2018, *128* (3), 422–442.

Fairlie, Robert, Alicia Robb, and David T Robinson, "Black and white: Access to capital among minority-owned start-ups," *Management Science*, 2022, *68* (4), 2377–2400.

Fieler, Ana Cecília and Sun Kyoung Lee, "Growing through Vintages," Working Paper 2026.

Ganguli, Ina, Jeffrey Lin, Vitaly Meursault, and Nicholas F. Reynolds, "Spreading Out Across Expanding Idea Space," Working Paper 35499, National Bureau of Economic Research July 2026.

Garfinkel, Jon A., Umang Khetan, and Amrita Nain, "Innovation Specificity," Working Paper, University of Iowa April 2023. Available at SSRN 4413754.

Gocmen, Ararat, Clara Martínez-Toledano, and Vrinda Mittal, "Private Capital Markets and Inequality," SSRN Working Paper 5166981, SSRN July 2025.

Gompers, Paul A., Will Gornall, Steven N. Kaplan, and Ilya A. Strebulaev, "How do venture capitalists make decisions?," *Journal of Financial Economics*, 2020, *135* (1), 169–190.

Gompers, Paul, Anna Kovner, Josh Lerner, and David Scharfstein, "Venture capital investment cycles: The impact of public markets," *Journal of Financial Economics*, 2008, *87* (1), 1–23.

González-Uribe, Juanita, "Exchanges of Innovation Resources inside Venture Capital Portfolios," *Journal of Financial Economics*, 2020, *135* (1), 144–168.

Gornall, Will and Ilya A Strebulaev, "The economic impact of venture capital: Evidence from public companies," *Available at SSRN 2681841*, 2021.

Hellmann, Thomas and Manju Puri, "Venture capital and the professionalization of start-up firms: Empirical evidence," *Journal of finance*, 2002, *57* (1), 169–197.

Hochberg, Yael V., Alexander Ljungqvist, and Yang Lu, "Whom You Know Matters: Venture Capital Networks and Investment Performance," *Journal of Finance*, 2007, *62* (1), 251–301.

Horton, John and Amir Sariri, "Marginal Ideas," Technical Report May 2026. Working Paper.

Kalyani, Aakash, "The Creativity Decline: Evidence from US Patents," Working Paper, Boston University July 2022. Available at SSRN 4318158.

Kaplan, Steven N and Per Strömberg, "Venture capitalists as pricipals: Contracting, screening, and monitoring," *The American Economic Review*, 2001, *91* (2), 426.

Karlsen, Johan, Katja Kisseleva, Aksel Mjøs, and David T Robinson, "Are Some Angels Better than Others?," Technical Report, National Bureau of Economic Research 2024.

Kerr, William R, Ramana Nanda, and Matthew Rhodes-Kropf, "Entrepreneurship as experimentation," *Journal of Economic Perspectives*, 2014, *28* (3), 25–48.

King, Robert G and Ross Levine, "Finance, entrepreneurship and growth," *Journal of Monetary Economics*, 1993, *32* (3), 513–542.

Kortum, Samuel and Josh Lerner, "Assessing the Contribution of Venture Capital to Innovation," *The RAND Journal of Economics*, 2000, *31* (4), 674–692.

Lerner, Josh and Ramana Nanda, "Venture Capital's Role in Financing Innovation: What We Know and How Much We Still Need to Learn," *Journal of Economic Perspectives*, 2020, *34* (3), 237–261.

Lyonnet, Victor and Léa H Stern, "Machine learning about venture capital choices," *Review of Financial Studies, Forthcoming*, 2026.

Ma, Song, "Technological obsolescence," *Review of Financial Studies*, 2025, p. hhaf059.

Manso, Gustavo, "Motivating innovation," *Journal of finance*, 2011, *66* (5), 1823–1860.

Nanda, Ramana and Matthew Rhodes-Kropf, "Investment cycles and startup innovation," *Journal of Financial Economics*, 2013, *110* (2), 403–418.

Paine, Fiona, "The Startup Valley of Death and Measuring Innovation," SSRN Working Paper 5521639, SSRN August 2025.

Parry, Charles, "Start-up Financing, Entry and Innovation," Technical Report, University of Cambridge December 2025.

Reimers, Nils and Iryna Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *Proceedings of EMNLP-IJCNLP*, 2019, pp. 3982–3992.

Robb, Alicia M and David T Robinson, "The capital structure decisions of new firms," *Review of Financial Studies*, 2014, *27* (1), 153–179.

Samila, Sampsa and Olav Sorenson, "Venture Capital, Entrepreneurship, and Economic Growth," *Review of Economics and Statistics*, 2011, *93* (1), 338–349.

Schumpeter, Joseph A, *Capitalism, Socialism and Democracy*, Allen & Unwin London, 1942.

Sørensen, Morten, "How smart is smart money? A two-sided matching model of venture capital," *Journal of Finance*, 2007, *62* (6), 2725–2762.

Tian, Xuan and Tracy Yue Wang, "Tolerance for failure and corporate innovation," *Review of Financial Studies*, 2014, *27* (1), 211–255.

Appendix

A Construction of Similarity Scores

The similarity score is constructed in three steps, described below. Steps 1 and 2 produce a firm-to-industry similarity score, capturing how a firm's description aligns with the typical discourse of its assigned industry category. Step 3 produces the pairwise firm-to-firm similarity score within each industry, which is used for classifying firms as novel or imitative.

A.1 Step 1a: TF-IDF Industry Vocabulary

For each industry category defined by `primaryindustrycode`, all firm descriptions belonging to the industry are pooled and "tokenized". Further, the stop words, tokens shorter than three characters, and purely numeric strings are removed. A TF-IDF (term frequency and inverse document frequency) weight is then computed for every word w in industry I, as follows:

$$\mathrm{TF}(w, I) = \frac{\text{count}(w \text{ in } I)}{\text{total words in } I}, \tag{12}$$

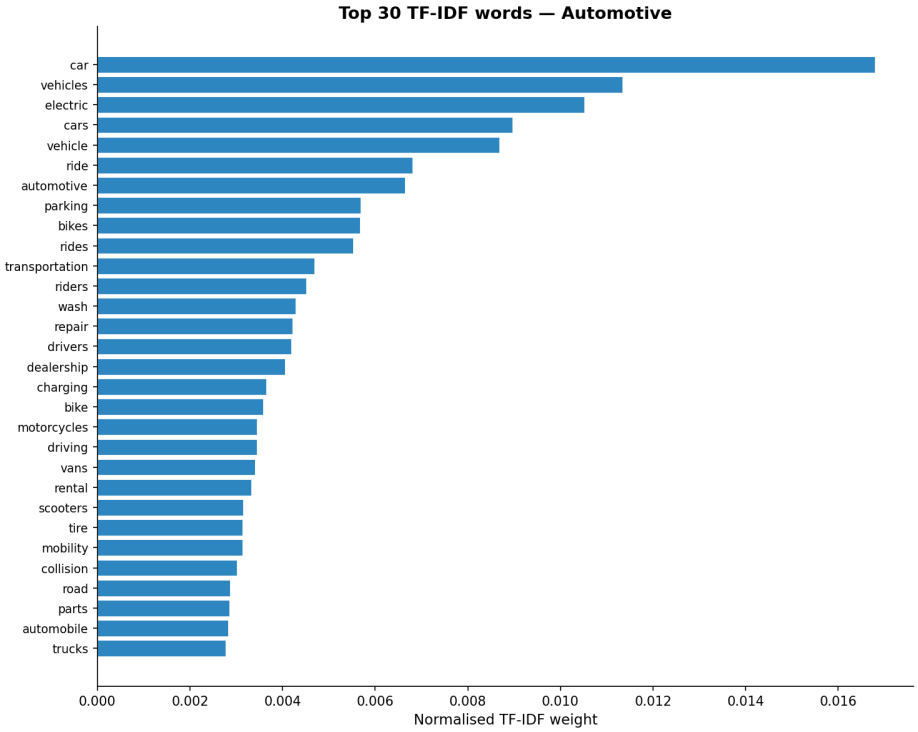
$$\mathrm{IDF}(w) = \log\left(\frac{N_{\text{industries}}}{1 + \#\{\text{industries containing } w\}}\right), \tag{13}$$

$$\text{weight}(w, I) = \mathrm{TF}(w, I) \times \mathrm{IDF}(w), \tag{14}$$

where $N_{industries}$ is the number of industries. Words with non-positive weight are removed, and each remaining weight is divided by the sum of all remaining weights in that industry, an ℓ_1 normalisation that makes the weights sum to one within each industry. The IDF component down-weights words that appear across many industries (e.g., services, solutions, platform), and up-weights words distinctive to a specific industry. The top 30 words by weight for each industry are retained as its vocabulary and forwarded to the next step.

As an illustration, Figure A1 shows the top-30 TF-IDF words for the Automotive industry. Terms such as "car", "vehicles", "electric", "ride", "bikes", and "riders" dominate, reflecting the heterogeneous mix of traditional automotive companies. Electric vehicle manufacturers and ride-sharing platforms are also classified under that category.

Figure A1: Top 30 TF-IDF Words in Automative Industry



Notes: This figure displays the top 30 TF-IDF words in the automative industry according to the Step 1a described above.

A.2 Step 1b: LLM-Generated Industry Reference Description

The top-30 TF-IDF word–weight pairs for each industry are submitted to Claude (`claude-opus-4-5`) via the Anthropic API. The model is prompted to synthesize those keywords into a concise 2–3 sentence description of what companies in the industry typically do, the products or services they offer, and their key characteristics. The prompt explicitly prohibits mention of TF-IDF terminology, ensuring that the output reads as natural prose. The generated description serves as the industry's canonical reference text in Step 2. Descriptions are cached to disk so that subsequent steps can be re-run without additional API calls.

A.3 Step 2: Firm-to-Industry Similarity

All firm descriptions and all industry reference descriptions are encoded into fixed-length dense vectors using "all-mpnet-base-v2", a Sentence Transformer (SBERT) model pre-trained on large corpora of natural language sentence pairs (Reimers and Gurevych, 2019). All embeddings are ℓ_2-normalised, so cosine

similarity reduces to the dot product. The firm-to-industry similarity score for firm i in industry I is:

$$s_i^{\text{ind}} = \mathbf{e}_i \cdot \mathbf{e}_I \ \in [0, 1], \tag{15}$$

where $\mathbf{e}_i$ is the unit-norm embedding of firm i's business description and $\mathbf{e}_I$ is the unit-norm embedding of the LLM-generated description for industry I. Similarity scores are restricted to the range between zero and one, with any negative value set to zero and any value above one set to one. This measure captures how well a firm's own language aligns with the typical discourse of its industry, and serves as a baseline indicator of industry conformity.

A.4 Step 3: Pairwise Firm-to-Firm Similarity

Within each industry, the same SBERT model is applied to encode all n firm descriptions into an $n \times d$ embedding matrix $\mathbf{E}$, from which the full pairwise cosine similarity matrix is computed as:

$$\mathbf{S} = \mathbf{E}\,\mathbf{E}^\top \ \in \mathbb{R}^{n \times n}. \tag{16}$$

Only the upper triangle of $\mathbf{S}$ is retained, storing each pair (i, j) exactly once. Each row of the output file corresponds to one firm pair and records the `pairwise_similarity` S_{ij} — the direct cosine similarity between their business descriptions — along with each firm's firm-to-industry score from Step 2 and their respective deal sizes.

A high value of S_{ij} indicates that firms i and j describe their businesses in semantically similar language, consistent with shared products, technologies, or business models. A newly incepted firm with consistently high pairwise similarity to incumbent peers is therefore classified as imitative; one with consistently low similarity is classified as novel.

B Model Proofs

Proof of Proposition 1 Let $I_j(A)$ be an indicator equal to one if VC j invests in a startup located in patch A, and zero otherwise. Under Assumption 1, each start-up receives investment from at most one VC, so

$$N(A) = \sum_{j \in \mathcal{J}} I_j(A).$$

Since VC j is active with probability π_j and, conditional on being active, invests in patch A with probability $P_j(A)$,

$$\Pr(I_j(A) = 1) = \pi_j P_j(A).$$

Therefore, by linearity of expectations,

$$\mathbb{E}[N(A)] = \sum_{j \in \mathcal{J}} \mathbb{E}[I_j(A)] = \sum_{j \in \mathcal{J}} \pi_j P_j(A).$$

Moreover, patch A receives no investment if and only if no VC invests in it. Under the assumed independence of VC activity and investment choices across investors,

$$\Pr(N(A) = 0) = \prod_{j \in \mathcal{J}} \Pr(I_j(A) = 0) = \prod_{j \in \mathcal{J}} \left[1 - \pi_j P_j(A)\right].$$

Hence,

$$\Pr(N(A) \geq 1) = 1 - \Pr(N(A) = 0) = 1 - \prod_{j \in \mathcal{J}} \left[1 - \pi_j P_j(A)\right].$$

■

Proof of Lemma 1. Comparing the above CDFs, we require

$$\frac{1}{n} \sum_{i=1}^{n} \exp\{q_{(i)} - \frac{\kappa}{b_j^\eta} d_{(i)j}\} \geq \frac{1}{I} \sum_{\iota \in \mathcal{I}} \exp\{q_{(\iota)} - \frac{\kappa}{b_j^\eta} d_{(\iota)j}\}, \quad \forall n$$

which is equivalent to

$$\frac{1}{n} \sum_{i=1}^{n} \exp\{q_{(i)} - \frac{\kappa}{b_j^\eta} d_{(i)j}\} \geq \frac{1}{I-n} \sum_{i=n+1}^{I} \exp\{q_{(i)} - \frac{\kappa}{b_j^\eta} d_{(i)j}\}, \quad \text{for} \quad n = 1, ..., I-1. \tag{17}$$

Substituting $q_{(i)} = \bar{q} + \alpha d_{(i)j}$ into (17), and cancelling the common factor $e^{\bar{q}}$, it is sufficient to show

that

$$\frac{1}{n}\sum_{i=1}^{n}\exp\left\{\left(\alpha-\frac{\kappa}{b_j^{\eta}}\right)d_{(i)j}\right\}\geq\frac{1}{I-n}\sum_{i=n+1}^{I}\exp\left\{\left(\alpha-\frac{\kappa}{b_j^{\eta}}\right)d_{(i)j}\right\}$$

for every $n=1,\ldots,I-1$.

Since

$$\alpha-\frac{\kappa}{b_j^{\eta}}<0$$

and start-ups are ordered by distance, the summands are weakly decreasing in i. Hence every summand on the left-hand side is weakly greater than every summand on the right-hand side. The inequality therefore holds for every n, which implies

$$F_j^A(d)\geq F_j^P(d)\qquad\text{for all } d.$$

■

Proof of Lemma 2. From equation (7), the distribution of absolute investment distances for investor j, evaluated at the n-th ordered distance, is

$$F_j^A(d_{(n)j})=\sum_{i=1}^{n}p_{(i)j}.$$

We claim that

$$F_s^A(d_{(n)})\geq F_g^A(d_{(n)})$$

for every n. By assumption, investors s and g face the same joint configuration of startup qualities and distances, so that $d_{(i)s}=d_{(i)g}=d_{(i)}$ and $q_{(i)s}=q_{(i)g}=q_{(i)}$.

The above statement is equivalent to

$$\sum_{i=1}^{n}\left(p_{(i)s}-p_{(i)g}\right)\geq 0$$

for every n. Noting that, for each startup (i),

$$p_{(i)s}=p_{(i)g}\frac{Z_g}{Z_s}\exp\left\{-\kappa\left(\frac{1}{b_s^{\eta}}-\frac{1}{b_g^{\eta}}\right)d_{(i)}\right\},$$

where

$$Z_j = \sum_{i \in \mathcal{I}} \exp\left\{ q_{(i)} - \frac{\kappa}{b_j^\eta} d_{(i)} \right\},$$

the condition is equivalent to

$$\sum_{i=1}^{n} p_{(i)g} \left[\frac{Z_g}{Z_s} \exp\left\{ -\kappa \left(\frac{1}{b_s^\eta} - \frac{1}{b_g^\eta} \right) d_{(i)} \right\} - 1 \right] \geq 0.$$

Define

$$h_i = \frac{Z_g}{Z_s} \exp\left\{ -\kappa \left(\frac{1}{b_s^\eta} - \frac{1}{b_g^\eta} \right) d_{(i)} \right\} - 1.$$

Since $b_s < b_g$, h_i is weakly decreasing in i. Moreover,

$$\sum_{i=1}^{I} p_{(i)g} h_i = \sum_{i=1}^{I} \left(p_{(i)s} - p_{(i)g} \right) = 0,$$

since $\sum_{i=1}^{I} p_{(i)j} = 1$ for each j. Because h_i is weakly decreasing and has weighted mean zero, there is a cutoff such that h_i is non-negative for startups closer than the cutoff and non-positive for startups farther away. If all startups are equidistant, then $h_i = 0$ for every i.

The partial sums

$$\sum_{i=1}^{n} p_{(i)g} h_i$$

therefore cannot become negative: once the negative terms begin, all remaining terms are non-positive, so a negative partial sum could not return to the terminal value of zero. Hence,

$$\sum_{i=1}^{n} p_{(i)g} h_i \geq 0$$

for every n, and therefore

$$F_s^A(d) \geq F_g^A(d)$$

for all d. ■

Proof of Corollary 1. Define $\psi_j = \kappa / b_j^\eta$ and note

$$\frac{\partial p_{ij}}{\partial \psi_j} = p_{ij} \left(\mathbb{E}_j^A[d_{ij}] - d_{ij} \right).$$

Then,

$$\frac{\partial E_j[d_{ij}]}{\partial \psi_j} = \sum_{i\in\mathcal{I}} d_{ij}\frac{\partial p_{ij}}{\partial \psi_j} = \sum_{i\in\mathcal{I}} d_{ij}p_{ij}\left(\mathbb{E}_j^A[d_{ij}] - d_{ij}\right)$$
$$= \mathbb{E}_j^A[d_{ij}]^2 - \mathbb{E}_j^A[d_{ij}^2] = -Var_j^A(d_{ij})$$

Finally,

$$\frac{\partial E_j[d_{ij}]}{\partial b_j} = \frac{\partial E_j[d_{ij}]}{\partial \psi_j}\frac{\partial \psi_j}{\partial b_j} = \eta\frac{\kappa}{b_j^{\eta+1}}Var_j^A(d_{ij}) > 0,$$

where the inequality follows from $Var_j^A(d_{ij}) > 0$. ■

Proof of Lemma 3 From equation (9), we require $\eta\frac{\kappa}{b_j^\eta}Var_j^A(d_{ij}) < \mathbb{E}_j^A[d_{ij}]$. For each VC j, the distribution of distances lies on $[0, d_{(I)j}]$. Then, the Bhatia–Davis inequality implies

$$Var_j^A(d_{ij}) \leq \mathbb{E}_j^A(d_{ij})[d_{(I)j} - \mathbb{E}_j^A(d_{ij})]$$

and, therefore, a sufficient condition for $\eta\frac{\kappa}{b_j^\eta}Var_j^A(d_{ij}) < \mathbb{E}_j^A[d_{ij}]$ is

$$\eta\frac{\kappa}{b_j^\eta}[d_{(I)j} - \mathbb{E}_j^A(d_{ij})] < 1.$$

Furthermore, a stronger but simpler sufficient condition is

$$\eta\frac{\kappa}{b_j^\eta}d_{(I)j} < 1,$$

which is equivalent to requiring that the elasticity of distance-related attractiveness with respect to portfolio breadth is less than one for every available start-up.[6] Since this expression is continuous and equals zero at $\eta = 0$, there exists some $\eta_j > 0$ for each investor j such that $\eta < \eta_j \implies \eta\frac{\kappa}{b_j^\eta}d_{(I)j} < 1$.

Therefore, defining $\bar{\eta} = \min_{j\in\mathcal{J}}\eta_j > 0$, $\eta < \bar{\eta}$ is a sufficient condition for expected relative distance to be decreasing in portfolio breadth for every investor $j \in \mathcal{J}$. ■

[6] Defining distance-related attractiveness as $\exp\{-(\kappa/b_j^\eta)d_{ij}\}$, its elasticity with respect to breadth is

$$\frac{\partial \log\exp\{-\frac{\kappa}{b_j^\eta}d_{ij}\}}{\partial \log b_j} = \eta\frac{\kappa}{b_j^\eta}d_{ij},$$

which is increasing in d_{ij} and maximized at $d_{(I)j}$ for each VC j.

Proof of Lemma 4 Follows directly from differentiating the expression in Proposition 1 with respect to δ.

C Estimation of Testable Implications

C.1 Construction of Idea Patches

The idea patches are constructed using the k-d tree approach. Each company is a point in the 768-dimensional embedding space. We build the tree using only the firms that existed before the shock we use (founded on or before 2013), and apply the same cuts to every company. Starting from these pre-shock firms, we repeat one operation until we reach a chosen depth. Every cut is fixed by three mechanical rules, described below.

1. **Which direction to cut.** Within the points currently in a region, we cut along the dimension of largest variance, that is, the direction in which those points are most spread out.
2. **Where to cut.** We place the cut at the median along that dimension, so the two resulting sub-regions hold the same number of pre-shock firms.
3. **When to stop.** Cutting until the tree reaches a preset depth L, which produces 2^L patches.

Because each region is split at its own median along its own most-spread direction, dense parts of the space are carved into many small patches while sparse parts stay large, and every patch ends up holding roughly the same number of pre-shock firms. We use a depth of 10, which gives 1,024 patches. The following figure illustrates the construction as the depth increases.

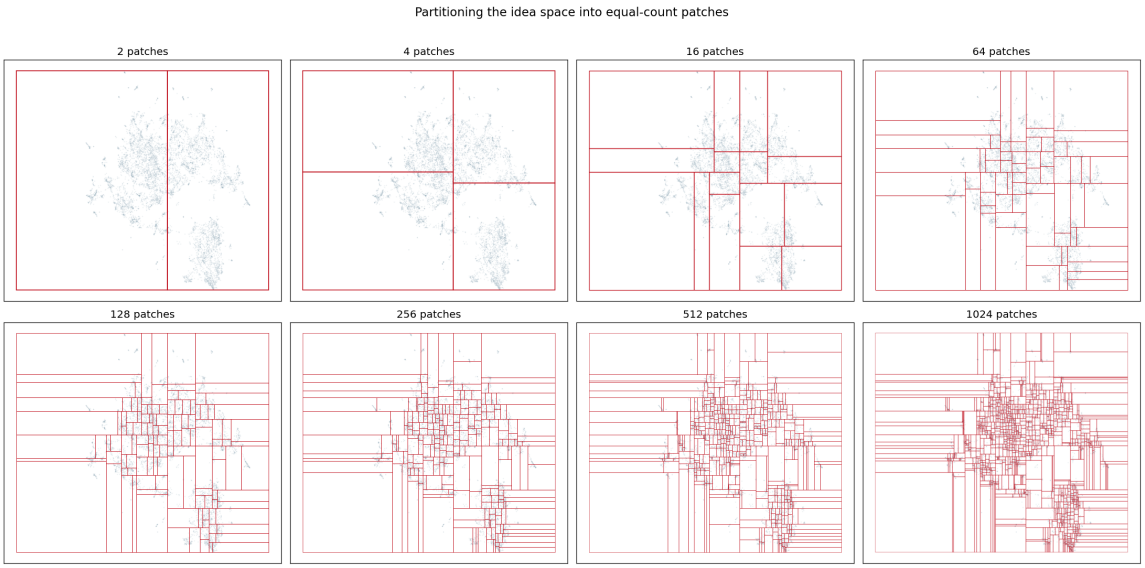


Table A1: k-d tree partition of the idea space

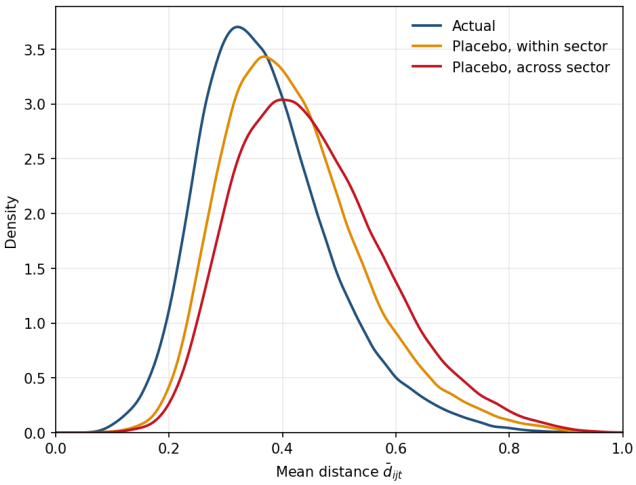
		Companies per patch					
Depth	Patches	Min	Median	Max	Empty (%)	Cohesion	vs. global
6	64	662	662	663	0.0	0.337	0.86
7	128	331	331	332	0.0	0.327	0.83
8	256	165	166	166	0.0	0.317	0.81
9	512	82	83	83	0.0	0.305	0.78
10	1024	41	41	42	0.0	0.290	0.74
11	2048	19	21	22	0.0	0.272	0.69
12	4096	9	10	11	0.0	0.246	0.63
13	8192	4	5	6	0.0	0.207	0.53
14	16384	2	3	3	0.0	0.145	0.37
15	32768	1	1	3	0.0	0.050	0.13

Notes: Partition built on pre-shock firms (founded ≤ 2013). Each split cuts the maximum-variance axis at its median, so patches hold an equal number of pre-shock firms; depth L yields 2^L patches. Cohesion is the mean within-patch cosine distance to the patch centroid, and "vs. global" scales it by the whole-space value of 0.392.

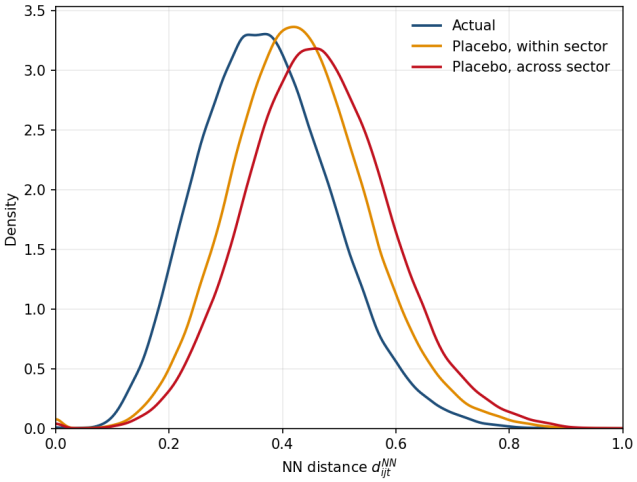
D Additional Figures and Tables

Figure A2: VC investments exhibit local bias (PDF)

(A) Mean distance

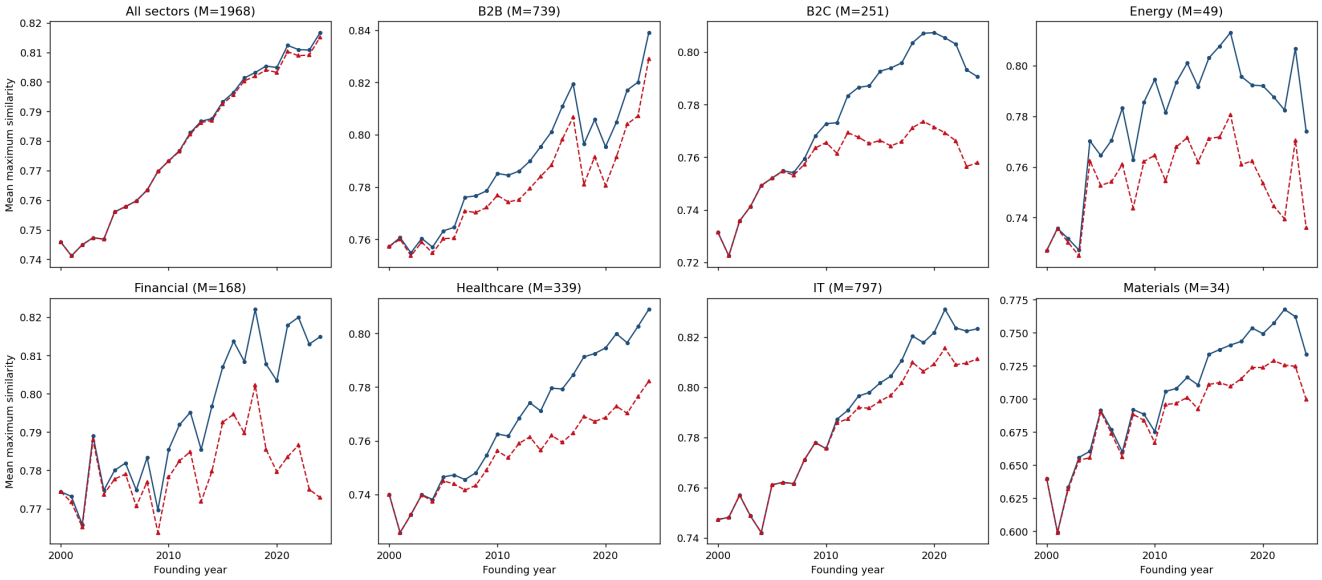


(B) NN distance



Notes: CDFs of the idea-space distance from a start-up to the prior five-year portfolio of the VC investing in it, in the new-company sample. Curves show the actual distance and two placebos that reassign the start-up to a random VC investing that year, within or across sector. Panel A uses the mean distance to the VC's prior five-year portfolio and Panel B the nearest-neighbor distance.

Figure A3: Similarity Scores For Fixed Number of Firms



Notes: This figure shows similarity scores over time. For each sector and founding year, we plot the average of each firm's highest similarity to same-sector firms founded in the previous ten years. The blue line uses all such earlier firms; the red line uses a random sample of at most M of them. M is set separately for each sector. It equals the largest number of earlier same-sector firms that any firm founded in 2000 can compare with, or the number of firms founded in that sector in 2000 if that is larger. Because M is at least as large as every 2000 firm's set of earlier firms, the two lines coincide exactly in 2000 and separate only later, as firms gain more than M earlier firms.