When Innovation Becomes Systematic: From Venture Capital Experimentation to Public-Market Risk

Caroline Genc* Stanislav Sokolinski†

July 2026

Abstract

We study when innovation news enters public equity markets as systematic rather than firm-specific risk. We show that this distinction depends on how venture capital is distributed across innovation directions. When startup investment spans distinct but connected directions, forming an innovation stack, innovation news is more likely to become systematic. When funding instead clusters within narrower directions, its effects remain more firm-specific. Using U.S. startup funding data from 2000 to 2024, we construct a cross-direction share that captures this contrast from Crunchbase industry category co-assignment. A higher cross-direction share predicts higher average pairwise correlation among S&P 500 stocks and a larger top-eigenvalue share of the implied correlation matrix over the following year, with the associated leading component aligned with the market. The effect is economically large, robust to macro-financial and innovation controls, and not explained by the overall similarity of the funded directions. Cross-sectional evidence further shows that the response is broad across firm pairs and firms rather than driven by a small set of highly exposed names. Venture capital experimentation therefore foreshadows whether innovation risk will remain firm-specific or become systematic.

*Broad College of Business, Michigan State University. Address: Department of Finance, 305 Eppley Center, East Lansing, MI 48824. E-mail: genccaro@msu.edu. Web: https://www.carolinegenc.com/.

†Broad College of Business, Michigan State University. Address: Department of Finance, 315 Eppley Center, East Lansing, MI 48824. E-mail: sokolins@msu.edu. Web: https://stanislavsokolinski.wordpress.com.

1 Introduction

Innovation news does not enter public equity markets in a single way. Some technological breakthroughs reprice only the firms most directly exposed, leaving the broader market largely unaffected. Others alter the value of many firms at once, generating common movements in stock returns and becoming a source of systematic risk. The distinction is fundamental to asset pricing: systematic innovation risk cannot be diversified away, whereas firm-specific innovation risk is largely diversifiable. Whether innovation news remains firm-specific or becomes systematic therefore shapes diversification opportunities, the factor structure of returns, and the transmission of technological change into public equity markets.[1]

Asset-pricing theory explains how innovation risk becomes systematic. In Pástor and Veronesi (2009), a new technology initially affects firms closest to the innovation, but its risk becomes systematic as adoption spreads through the economy. Yet by the time this transition becomes visible in public prices, it is already underway. We show that the emergence of systematic innovation risk can be detected earlier through signals from the venture capital market.

The venture capital market provides a natural setting in which to look for such signals because new technologies are financed there before the resulting firms, products, and economic linkages are fully reflected in public valuations. Venture funding reveals not only how much innovation is funded, but also how experimentation is organized. In some periods, funding is concentrated within a narrow innovation direction. In others, it spans distinct but connected directions, supporting experimentation across complementary parts of a broader technological ecosystem. We capture this organization with a single measure, *the cross-direction share*, and show that it forecasts public-market comovement and factor structure one year ahead.

When funding spans distinct but connected directions, innovation news is more likely to contain a common component that reaches many firms. Public stocks subsequently move more together and the market becomes more one-factor-like. When funding instead concentrates within narrower directions, innovation news remains more firm-specific and comovement recedes. What predicts future comovement is therefore not the amount of venture activity, but how that activity is distributed across connected innovation directions.

We build the measure from U.S. startup funding between 2000 and 2024. Crunchbase assigns startups to fine-grained, non-exclusive industry categories, which we treat as innovation directions. Because startups can be assigned to several categories, category co-assignment reveals which directions are empirically related: two directions are close when many of the

[1] Average pairwise correlation is priced in equity-index options (Driessen et al., 2009) and forecasts market returns (Pollet and Wilson, 2010), and the common component of firm-level volatility commands its own premium (Herskovic et al., 2016).

same startups combine them.[2] The resulting statistic has a simple interpretation. Draw two venture dollars at random. They can be related because they fund the same direction, or because they fund different but connected directions. The cross-direction share s_t is the fraction of total relatedness that comes from the second source:

$$s_t = \frac{\text{relatedness from different but connected directions}}{\text{total relatedness among venture dollars}}.$$

A higher value means more relatedness comes from funding distinct but connected directions; a lower value means relatedness comes primarily from within-direction concentration. Thus s_t measures the composition of total relatedness, not total funding or total relatedness itself.

A simple model links venture composition to the factor structure of public equity returns and delivers testable asset-pricing predictions. Each firm's unexpected return combines non-innovation news with innovation news. The innovation component depends on a firm-level exposure, which can be positive or negative: some firms benefit from a connected innovation stack, while others are threatened by it. The key link is that venture composition determines how common innovation news is across firms. In the model, s_t captures the cross-firm correlation of innovation news, so the resulting pairwise return correlation between firms i and j is approximately

$$\text{future return correlation}_{ij} \;\approx\; s_t \times \text{innovation exposure}_i \times \text{innovation exposure}_j.$$

This expression delivers three implications under broadly same-signed exposures, when the firms lifted by innovation outweigh those it displaces. First, a higher cross-direction share should forecast higher average pairwise correlation. Second, it should forecast a larger share of correlation-matrix variation absorbed by the leading component, provided that component is market-aligned, as it is in the data. Third, the effect should be broad. Because each pair's sensitivity is the product of the two firms' exposures, the response should show up across the correlation matrix—with most pairs and most firms moving in the same direction—rather than in a few names. The model also gives a sharp placebo: total relatedness does not enter the correlation structure, so composition should predict comovement, while the level should not.

The data line up with these predictions. In monthly regressions, a one-standard-deviation increase in the cross-direction share predicts higher average pairwise correlation among S&P 500 stocks over the next year. The effect is economically large, ranging from approximately 0.36 to 0.63 standard deviations of the outcome across the main specifications.

[2] The construction recovers technological proximity from revealed choices rather than manual classification, in the spirit of Bloom et al. (2013), Hoberg and Phillips (2016), and Kelly et al. (2021).

The top-eigenvalue share of the implied correlation matrix moves in the same direction: when the cross-direction share is higher, public returns become more organized around a common component. The associated leading component is market-aligned, so comovement concentrates in the market direction rather than in a separate cluster. By contrast, total relatedness has no predictive power once composition is held fixed. The signal is strongest at the twelve-month horizon and remains stable after controlling for uncertainty, funding conditions, investor sentiment, and direct measures of innovation activity.

The same pattern appears inside the correlation matrix. This cross-sectional anatomy is important because the aggregate results could, in principle, come from a narrow set of exposed firms or a few unusually exposed pairs. That is not what we find. In pair-by-pair regressions, 77.6 percent of firm pairs respond positively to the cross-direction share. When those pair estimates are aggregated to the firm level, 84.9 percent of firms contribute positively. Positive contribution mass is distributed broadly rather than concentrated in a small set of pairs or firms. Average pairwise correlation, factor structure, pair responses, and firm contributions therefore line up with a single set of innovation exposures—the over-identifying discipline at the heart of the model.

We interpret these estimates as forecasts, not as the effect of an experiment that randomly reallocates venture dollars. Venture composition is an equilibrium object, but that is precisely why it is informative: it summarizes how investors, founders, and technologies are organizing experimentation before the resulting risks are fully visible in public prices. The empirical design shows that this upstream signal contains information beyond the usual candidates. The relation survives controls for uncertainty, financing activity, and investor sentiment, so it is not simply a reflection of macro-financial conditions. It also survives direct innovation proxies—patent value, product-market fluidity, and technology news attention—and controls for funding flows into major technology areas, so it is not only capturing innovation intensity or which sectors are hot. The sharpest distinction is within the venture measure itself: total relatedness carries no signal once composition is held fixed. What remains is a compositional signal—not how much venture capital is deployed, and not simply where it goes, but how it is organized across related directions.

Our first contribution is to the literature on return comovement. Existing work explains comovement through forces inside public markets: common fundamentals and market-wide volatility (Pindyck and Rotemberg, 1993; Campbell et al., 2001), information environments (Morck et al., 2000; Veldkamp, 2006), investor categories and habitats (Barberis et al., 2005), and time-varying correlation risk (Engle and Kelly, 2012). A second strand emphasizes real linkages among listed firms, including granular size distributions (Gabaix, 2011) and production networks (Acemoglu et al., 2012; Herskovic, 2018). We move the source of comovement upstream. Rather than starting from trading frictions, contemporaneous network shocks, or characteristics of listed firms, we study how risks that originate in private innovation markets later appear as common components of public returns.

The paper also contributes to the literature connecting innovation and asset prices. That literature studies how technological revolutions, patents, and displacement risk affect valuations, risk premia, and growth (Hobijn and Jovanovic, 2001; Gârleanu et al., 2012; Kogan and Papanikolaou, 2014; Kogan et al., 2017). We focus instead on the covariance dimension of innovation news. In Pástor and Veronesi (2009), innovation risk becomes systematic as adoption spreads, and the transition is inferred from public prices as it unfolds. We locate an earlier stage of that transition, before the relevant firms and technologies are fully incorporated into public-market valuations.

Finally, we contribute to the literature on venture capital and innovation, which studies how VC affects entrepreneurship, the rate of innovation, and growth (Kortum and Lerner, 2000; Samila and Sorenson, 2011; Nanda and Rhodes-Kropf, 2013; Lerner and Nanda, 2020). We shift attention from the quantity of venture capital to its architecture. The relevant object is not only how much experimentation private investors finance, but how that experimentation is organized across innovation directions. This perspective links private innovation finance to asset-pricing outcomes that are usually studied only after firms enter public markets.

2 Conceptual Framework

2.1 Venture allocation and composition

Two venture dollars can be technologically related for two different reasons. They can fund the same innovation direction, or they can fund distinct directions that are themselves connected. The first source is *within-direction concentration*; the second is *cross-direction relatedness*. The distinction is about the composition of total relatedness, separate from both the amount of venture capital deployed and total relatedness.

Suppose there are K innovation directions. At date t, venture capital is allocated across them according to the share vector $p_t = (p_{1t}, \dots, p_{Kt})'$, with $p_{kt} \geq 0$ and $\sum_k p_{kt} = 1$. A symmetric matrix W_t records similarity across directions, with $W_{kk,t} = 1$ and $W_{k\ell,t} \in [0, 1]$ for $k \neq \ell$. Let $M_t \equiv W_t - I$ retain only the off-diagonal similarities. Section 3.3 constructs these objects from startup funding and category co-assignment. The two components of relatedness are

$$R_t \equiv p_t' p_t = \sum_k p_{kt}^2, \qquad S_t \equiv p_t' M_t p_t = \sum_{k \neq \ell} p_{kt} p_{\ell t} W_{k\ell,t}. \tag{1}$$

If two venture dollars are drawn independently from p_t, R_t is the probability that they fund the same direction, while S_t is the similarity-weighted relatedness generated when they fund different directions. Their sum is total relatedness,

$$\alpha_t \equiv R_t + S_t = p_t' W_t p_t. \tag{2}$$

We summarize its composition with the *cross-direction share*

$$s_t \equiv \frac{S_t}{S_t + R_t} \in [0, 1). \tag{3}$$

Thus, s_t is the fraction of total relatedness that comes from funding across distinct but connected directions. A higher value indicates broader funding across a connected innovation stack; a lower value indicates that relatedness comes primarily from repeated funding within the same directions. Section 3.3 develops the probabilistic interpretation and empirical construction in detail.

2.2 Venture composition and unexpected returns

We link venture composition at date t to the factor structure of public equity returns measured at a later horizon. Let $\tilde{r}_{i,t}^{(h)}$ denote firm i's unexpected return over the public-market return window ending at $t+h$, where h indexes the forecast horizon. Define the corresponding conditional correlation between firms i and j as

$$\rho_{ij,t}^{(h)} \equiv \mathrm{Corr}_t \left(\tilde{r}_{i,t}^{(h)}, \tilde{r}_{j,t}^{(h)} \right). \tag{4}$$

For ease of exposition, we write the framework for returns scaled by baseline volatility,

$$x_{i,t}^{(h)} \equiv \frac{\tilde{r}_{i,t}^{(h)}}{\sigma_i^0}, \tag{5}$$

where σ_i^0 is firm i's baseline volatility. The scaling does not require transforming the data: because correlation is scale-free, $\mathrm{Corr}(\tilde{r}_{i,t}^{(h)}, \tilde{r}_{j,t}^{(h)}) = \mathrm{Corr}(x_{i,t}^{(h)}, x_{j,t}^{(h)})$, so the empirical work estimates correlations from raw returns directly.

We model each firm's scaled unexpected return as a mix of two kinds of news: non-innovation news, which moves the firm for reasons outside the venture-composition channel, and innovation news, which the channel transmits,

$$x_{i,t}^{(h)} = \sqrt{1-a_i^2} \underbrace{e_{i,t}^{(h)}}_{\text{non-innovation news}} + a_i \underbrace{g_{i,t}^{(h)}}_{\text{innovation news}} . \tag{6}$$

Both shocks are mean-zero with unit variance, so the squred loadings sum to one: a_i is the firm's exposure to innovation and a_i^2 the share of its return variance that innovation drives. The exposure can be positive or negative—a firm may ride an innovation wave or be displaced by it. We place no restriction on how the non-innovation shocks comove: their cross-firm correlations, $\mathrm{Corr}(e_{i,t}^{(h)}, e_{j,t}^{(h)}) = \tilde{\rho}_{ij,t}^{0,(h)}$, are left free. Scaled by the non-innovation loadings, they produce the baseline return comovement $\rho_{ij,t}^{0,(h)} \equiv \sqrt{1-a_i^2}\sqrt{1-a_j^2}\,\tilde{\rho}_{ij,t}^{0,(h)}$, the correlation between firms i and j that would prevail absent the venture-composition channel.

Innovation news, in turn, splits into a component common to all firms and a firm-specific one:

$$g_{i,t}^{(h)} = \sqrt{s_t}\, f_t^{(h)} + \sqrt{1-s_t}\, u_{i,t}^{(h)}, \tag{7}$$

where $f_t^{(h)}$ is common across firms and $u_{i,t}^{(h)}$ is specific to firm i, both orthogonal and standardized. The firm-specific pieces are uncorrelated across firms, $\mathrm{Corr}(u_{i,t}^{(h)}, u_{j,t}^{(h)}) = 0$ for $i \neq j$, so purely local innovation news moves no two firms together: the common component $f_t^{(h)}$ is the only source of innovation comovement.

The cross-direction share s_t is the weight on the common component, making it the cross-firm correlation of innovation news, $\mathrm{Corr}(g_{i,t}^{(h)}, g_{j,t}^{(h)}) = s_t$. Composition and exposure therefore play distinct roles: s_t determines how common innovation news is, while a_i determines how firm i responds. This equation is the framework's economic mapping, not the empirical construction of s_t; Section 3.3 builds the measure exclusively from startup funding and category assignments.

The mapping reflects technological complementarity. Funding across connected directions supports innovations whose payoffs depend on progress elsewhere in the innovation stack, a central feature of general-purpose technologies (Bresnahan and Trajtenberg, 1995). Their news can therefore reach many public firms at once. Funding concentrated within a narrower direction may still produce important innovation, but more of its valuation effect remains firm-specific and diversifiable (Acemoglu et al., 2012). In this sense, s_t is a predictive state variable for the transition from idiosyncratic to systematic innovation risk described by Pástor and Veronesi (2009); the empirical analysis does not require interpreting venture allocation as a causal treatment.

2.3 Empirical implications

Taking the correlation of $x_{i,t}^{(h)}$ and $x_{j,t}^{(h)}$ for $i \neq j$ gives

$$\rho_{ij,t}^{(h)} = \rho_{ij,t}^{0,(h)} + s_t a_i a_j, \tag{8}$$

where the non-innovation terms contribute the baseline $\rho_{ij,t}^{0,(h)}$, the common shock contributes $s_t a_i a_j$, and the firm-specific shocks drop out. A pair's correlation moves with venture composition through the cross-direction share s_t, scaled by the exposure product $a_i a_j$.[3]

Averaging (8) across pairs, the effect of a higher cross-direction share on average pairwise correlation has the sign of $\sum_{i \neq j} a_i a_j = (\sum_i a_i)^2 - \sum_i a_i^2$, the squared net exposure of the market less the dispersion of exposures across firms. The sum is positive when firms' responses to innovation line up more than they offset, so the firms that gain from an innovation outweigh those it displaces. We call this configuration *broadly same-signed* exposures: the common shock pushes most firms the same way, and a higher s_t raises average comovement. Implication 1 summarizes this intuition. Proofs and derivative expressions are in Appendix A.[4]

Implication 1 (Average pairwise correlation)**.** *Under broadly same-signed exposures, a higher cross-direction share forecasts higher average pairwise return correlations.*

Higher average pairwise correlation says nothing on its own about the structure of comovement. The same rise could come from a new factor—a cluster of firms comoving among themselves but apart from the market—lifting correlations while leaving the market no more dominant than before (Kogan and Papanikolaou, 2014). When same-signed exposure products dominate and the leading eigenvector is already close to the market direction, the added covariance instead strengthens that market-aligned component. The top-eigenvalue share then rises along with average correlation. Implication 2 separates this market-wide response from an increase in comovement confined to a narrower cluster. Section 4 evaluates the maintained market-alignment condition and the resulting prediction.

Implication 2 (Factor structure)**.** *Under broadly same-signed exposures, a higher cross-direction share forecasts a higher top-eigenvalue share of the implied correlation matrix, if the first principal component of returns is close to the market direction.*

[3] The implied matrix is a valid correlation matrix whenever the residual covariance matrix—diagonal $1-a_i^2$, off-diagonal $\rho_{ij,t}^{0,(h)}$—is positive semidefinite, equivalently when the baseline correlation matrix dominates $\mathrm{diag}(a_i^2)$.

[4] Taking these comparative statics to the data requires a conditional independence assumption: the exposures a_i must not mechanically track which broad technology areas are funded, total relatedness α_t, or baseline correlations driven by volatility or business-cycle regimes. We treat this as the identifying assumption and return to it with technology-area and macro controls in the empirical section.

Implications 1 and 2 concern aggregate comovement and factor structure. Equation (8) also produces cross-sectional restrictions. A pair responds positively when its two firms have same-signed exposures and negatively when one benefits from the innovation while the other is displaced. Because the same exposure a_i enters every pair containing firm i, the pair coefficients can also be aggregated into firm-level contributions. Under the broadly same-signed configuration, positive pair responses and positive firm contributions should predominate. Implication 3 states these restrictions; Section 4 examines their distribution and concentration.

Implication 3 (Cross section of correlations)**.** *Under broadly same-signed exposures, a higher cross-direction share has the following effects.*

(a) It shifts each pair's correlation by the product of the two firms' exposures,

$$\beta_{ij}^{(h)} \equiv \frac{\partial \rho_{ij,t}^{(h)}}{\partial s_t} = a_i a_j, \qquad i \neq j, \tag{9}$$

so most pairs respond positively.

(b) Each firm's contribution to the summed pairwise effect $\sum_{i<j} \beta_{ij}^{(h)}$, splitting every pair equally between its two firms,

$$k_i^{(h)} \equiv \frac{1}{2} \sum_{j \neq i} \beta_{ij}^{(h)} = \frac{1}{2}\, a_i \sum_{j \neq i} a_j, \tag{10}$$

inherits the sign of its exposure a_i, so most firms contribute positively. Both patterns hold broadly across the cross-section rather than through a few firms.

3 Data and Measurement

3.1 Data sources

Our venture capital data come from Crunchbase, which reports startup funding rounds, financing dates and amounts, firm characteristics, and industry categories. Startups may be assigned to more than one category. This non-exclusive structure is central to our design because a startup's set of categories reveals the innovation directions that the firm combines. The funding data tell us where venture capital is allocated, while the category assignments provide the information needed to recover relatedness across those directions.

We retain U.S. startups with funding rounds announced between 2000 and 2024. We classify a round as venture capital financing when Crunchbase labels it as "seed" or its investment type contains the term "series." This rule captures early-stage and staged equity financing while excluding funding events that do not fit the venture capital setting. We remove observations with missing category information because they cannot be mapped into the innovation-direction space. We also exclude firms with inconsistent founding and financing dates.

Funding is allocated to categories at the round level. When a startup belongs to a single category, the full round amount is assigned to that direction. When it belongs to several categories, we divide the funding amount equally across them. This fractional assignment ensures that each round contributes exactly its observed funding total to the aggregate allocation. It therefore avoids mechanically multiplying funding for multi-category startups while preserving the information that these firms operate across several directions.

Crunchbase provides both granular industry categories and broader category groups. Our baseline uses the granular taxonomy because it preserves the detail needed to separate investment concentrated within a direction from investment spread across distinct but related directions.[5] In our sample, the taxonomy contains 801 categories, including Artificial Intelligence, Machine Learning, SaaS, Biotechnology, Medical Devices, Marketplace, and Semiconductors. Their category co-assignment across startups provides the raw input to the category-similarity matrix defined below.

Public-market outcomes combine CRSP daily returns with historical S&P 500 constituent data. We calculate daily excess returns using the risk-free rate from Kenneth French's Data Library. For each formation month, eligible securities are common stocks that belonged to the S&P 500 at some point during the trailing 12-month return window, had a price of at least $5 at formation, and had complete daily return coverage throughout the window. Using historical membership allows the return universe to evolve with the index rather than imposing the current constituent list on earlier periods. Complete return coverage produces a balanced panel within each formation window, which is important for estimating a common covariance matrix. When multiple securities map to the same firm, we retain the security with the largest market capitalization.

The controls address four potential sources of common variation. First, aggregate uncertainty may raise correlations independently of innovation. We measure policy uncertainty with the Economic Policy Uncertainty index (Baker et al., 2016) and market uncertainty with the monthly average Cboe VIX. Second, broad financing conditions may jointly affect venture capital allocation and public markets. We capture these conditions using gross U.S.

[5] Broader groups combine categories that may play different roles in entrepreneurial experimentation. Moreover, because granular categories can map into multiple groups, aggregation weakens the category co-assignment variation that identifies relatedness across directions.

IPO issuance from Jay Ritter's IPO database and total U.S. venture funding constructed from Crunchbase rounds. Third, we control for investor sentiment using the orthogonalized sentiment index of Baker and Wurgler (2006).

Finally, we include alternative measures of innovation and product-market change to separate the composition of venture activity from broader shifts in the innovation environment. These controls are value-weighted TNIC breadth (Hoberg and Phillips, 2016), value-weighted product-market fluidity (Hoberg et al., 2014), aggregate patent value (Kogan et al., 2017), and BNBC News Attention (Tech Buzz) for technology-related topics (Bybee et al., 2024). Together, they capture changes in product-market structure, competitive movement, the economic value of patented innovation, and the amount of business-news attention devoted to technology.

3.2 Estimating public-market comovement

We construct two monthly measures of public-market comovement from daily excess returns on eligible S&P 500 firms: average pairwise correlation and the top-eigenvalue share of the implied correlation matrix. Both outcomes are constructed from the same covariance estimate and summarize complementary features of the return structure.

For each formation month t, we estimate the covariance matrix using returns from the preceding 12 months. Returns are demeaned within the window. Because the cross-section contains many firms relative to the number of daily observations, the sample covariance matrix can be noisy. We therefore apply Ledoit–Wolf shrinkage toward a scaled identity target (Ledoit and Wolf, 2004). The procedure regularizes the covariance estimate and limits the influence of extreme sample covariances without imposing the factor or industry structure that the outcome variables are intended to measure.

Let Σ_t denote the resulting shrunk covariance matrix, and let N_t denote the number of firms in the formation-month universe. We convert Σ_t into the implied correlation matrix $\mathcal{R}_t$, whose elements are

$$\mathcal{R}_{ij,t} = \frac{\Sigma_{ij,t}}{\sqrt{\Sigma_{ii,t}\Sigma_{jj,t}}}.$$

Our first outcome, average pairwise correlation, is the mean of the off-diagonal elements of $\mathcal{R}_t$:

$$\bar{\rho}_t = \frac{1}{N_t(N_t-1)} \sum_{i \neq j} \mathcal{R}_{ij,t}.$$

This statistic summarizes the comovement of a typical pair of public firms and provides the direct empirical counterpart to the model's average-pairwise-correlation prediction in Section 2. It averages the bilateral relations across the entire correlation matrix without imposing a particular factor structure.

To examine whether the aggregate relation is broad or concentrated in a small part of the market, we also retain each off-diagonal element as a pair-level outcome:

$$\rho_{ij,t} = \mathcal{R}_{ij,t}, \qquad i \neq j.$$

These pair-level correlations allow us to determine whether changes in average comovement reflect a widespread shift across firm pairs or unusually large responses among a narrow set of firms.

Our second outcome is the top-eigenvalue share of the implied correlation matrix:

$$\Lambda_t = \frac{\lambda_1(\mathcal{R}_t)}{\mathrm{tr}(\mathcal{R}_t)} = \frac{\lambda_1(\mathcal{R}_t)}{N_t},$$

where $\lambda_1(\mathcal{R}_t)$ is the largest eigenvalue of $\mathcal{R}_t$. The eigenvalues decompose total standardized return variance across orthogonal principal components, so Λ_t is the fraction associated with the leading component. Because the trace of a correlation matrix equals the number of firms, the denominator is N_t.

The economic interpretation is direct. A high value of Λ_t means that a large fraction of standardized return variation is concentrated in one common direction, so the market is closer to a one-factor structure. A low value means that return variation is spread across several components and is therefore less dominated by a single source of comovement. The measure captures the strength of the leading component, but it does not assign that component an economic identity; we examine its market alignment separately. Taken together, average pairwise correlation, pair-level correlations, and the top-eigenvalue share describe the average strength, cross-sectional breadth, and factor concentration of public-market comovement.

3.3 Measuring venture composition

We next construct empirical measures of how venture capital is allocated across innovation directions. The construction maps the objects in the model into observable startup data. Crunchbase industry categories represent innovation directions, funding amounts determine the allocation of venture capital across those directions, and category co-assignment reveals which directions are empirically related. This approach follows the broader idea of recovering economic proximity from observed product descriptions, patents, or market links (Bloom et al., 2013; Hoberg and Phillips, 2016; Kelly et al., 2021). Our setting differs in that the proximity space is estimated among private startups and then used to predict the covariance structure of public firms.

The construction has three steps. First, we aggregate funding rounds to measure the monthly allocation of venture capital across Crunchbase categories. Second, we estimate a time-varying category-similarity matrix from category co-assignment across startups. Third, we combine the allocation shares with this matrix to separate within-direction concentration from cross-direction relatedness.

For each month t, we aggregate all funding rounds announced during the preceding 12 months.[6] Let $\mathcal{F}_t$ denote the set of rounds in this window. Using the fractional assignment rule described above, a round $f \in \mathcal{F}_t$ with amount A_f and category set I_f contributes $A_f/|I_f|$ dollars to each category in that set. Total funding assigned to category k is

$$D_{kt} = \sum_{f \in \mathcal{F}_t} \frac{A_f}{|I_f|} \mathbf{1}\{k \in I_f\},$$

and its share of venture funding is

$$p_{kt} = \frac{D_{kt}}{\sum_{j=1}^{K} D_{jt}}, \qquad \sum_{k=1}^{K} p_{kt} = 1.$$

The vector p_t therefore describes the distribution of venture capital across innovation directions, abstracting from the total funding in the window. A larger p_{kt} means that category k receives a larger funding share.

We next estimate how closely the directions are related. For each month t and category k, let U_{kt} denote the set of startups assigned to category k during the baseline trailing 12-month similarity-estimation window. For two distinct categories k and ℓ, we measure similarity using the Jaccard index:

$$W_{k\ell,t} = \frac{|U_{kt} \cap U_{\ell t}|}{|U_{kt} \cup U_{\ell t}|}, \qquad k \neq \ell,$$

with $W_{kk,t} = 1$. The index compares the number of startups assigned to both categories with the number assigned to either category. It equals zero when the two startup sets do not overlap and approaches one as their assignments become more similar. Normalizing by the union prevents large categories from appearing close merely because they contain many startups. Let W_t denote the resulting symmetric similarity matrix, and define $M_t \equiv W_t - I$ to retain only similarity across distinct directions.

We combine the funding shares and the similarity matrix to measure the relatedness of the venture capital allocation. Total similarity-weighted relatedness, which we refer to below

[6] The 12-month window smooths idiosyncratic round timing while keeping the allocation measure close to the current organization of private-market experimentation.

as total relatedness, is

$$\alpha_t = p_t' W_t p_t = p_t' p_t + p_t' M_t p_t.$$

This quadratic form has a useful interpretation. Draw two venture dollars independently from the allocation p_t and compare their assigned directions. Then α_t is their expected similarity, with identical directions receiving a value of one and distinct directions receiving the corresponding entry of W_t. The decomposition separates the contributions from same-direction and cross-direction draws.

The contribution from same-direction draws is within-direction concentration:

$$R_t \equiv p_t' p_t = \sum_k p_{kt}^2,$$

Because $W_{kk,t} = 1$, R_t is the probability that the two dollars are assigned to the same direction. It is also the Herfindahl concentration of the funding shares, so it rises when venture capital is concentrated in a smaller set of categories. The contribution from draws assigned to distinct directions is cross-direction relatedness:

$$S_t \equiv p_t' M_t p_t = \sum_{k \neq \ell} p_{kt} p_{\ell t} W_{k\ell,t}.$$

S_t is the expected similarity contributed by draws assigned to different directions: each pair k, ℓ is weighted by $W_{k\ell,t}$. Closely connected directions therefore receive more weight, while unrelated directions contribute little or nothing. By construction, total relatedness is $\alpha_t = R_t + S_t$.

Our main measure is the *cross-direction share*, defined as

$$s_t \equiv \frac{S_t}{R_t + S_t} = \frac{S_t}{\alpha_t}.$$

The cross-direction share measures the composition of total relatedness rather than its level. A high s_t means that a large fraction of relatedness comes from funding across distinct but connected directions; a low s_t means that relatedness comes primarily from within-direction concentration. Equivalently, s_t is the fraction of expected similarity attributable to draws assigned to different directions.

Numerical illustration. Figure 1 illustrates the construction in a simple economy with six startups and three innovation directions: Artificial Intelligence, Machine Learning, and Software. The example separates the two inputs to the measure—the relatedness between directions and the allocation of venture capital across them—and then shows how they combine to produce R_t, S_t, and s_t.

Panel (a) reports the startup–category assignments. The sets of startups assigned to each category are

$$U_{\text{AI}} = \{s_1, s_2, s_4, s_5\}, \qquad U_{\text{ML}} = \{s_1, s_2, s_3\}, \qquad U_{\text{Soft.}} = \{s_3, s_5, s_6\}.$$

AI and Machine Learning overlap through startups s_1 and s_2, AI and Software overlap through startup s_5, and Machine Learning and Software overlap through startup s_3. Dividing each intersection by the corresponding union gives the Jaccard similarities

$$\begin{aligned} W_{\text{AI,ML}} &= \frac{|\{s_1, s_2\}|}{|\{s_1, s_2, s_3, s_4, s_5\}|} = \frac{2}{5} = 0.40, \\ W_{\text{AI,Soft.}} &= \frac{|\{s_5\}|}{|\{s_1, s_2, s_3, s_4, s_5, s_6\}|} = \frac{1}{6} \simeq 0.17, \\ W_{\text{ML,Soft.}} &= \frac{|\{s_3\}|}{|\{s_1, s_2, s_3, s_5, s_6\}|} = \frac{1}{5} = 0.20. \end{aligned}$$

Panel (b) collects these values in the similarity matrix W. The diagonal entries equal one because each direction is perfectly similar to itself. The off-diagonal entries show that AI and Machine Learning are the closest pair in the example, while the other two pairs are related more weakly.

Panel (c) adds the venture capital allocation vector $p_t = (0.40, 0.35, 0.25)'$, so AI receives 40 percent of funding, Machine Learning receives 35 percent, and Software receives 25 percent. Combining p_t with W produces the contribution matrix C_t, whose elements are

$$C_{k\ell,t} = p_{kt} p_{\ell t} W_{k\ell,t}.$$

Each cell is the similarity-weighted relatedness generated by the ordered pair of directions (k, ℓ). The diagonal cells correspond to two venture dollars assigned to the same direction and therefore sum to within-direction concentration:

$$R_t = 0.1600 + 0.1225 + 0.0625 = 0.3450.$$

The off-diagonal cells correspond to dollars assigned to different directions. Because the matrix is symmetric, each unordered direction pair appears twice. Using the rounded entries displayed in the figure, cross-direction relatedness is

$$S_t \simeq 2(0.0560 + 0.0170 + 0.0175) = 0.1810.$$

Total similarity-weighted relatedness is therefore

$$\alpha_t = R_t + S_t \simeq 0.3450 + 0.1810 = 0.5260,$$

and the main composition measure is

$$s_t = \frac{S_t}{R_t + S_t} = \frac{0.1810}{0.5260} \simeq 0.344.$$

Thus, approximately 34.4 percent of total relatedness comes from funding across distinct but connected directions. The example shows that two venture capital allocations can have the same α_t but different s_t because the sources of their relatedness differ.

Cross-direction share in the data. We next use three figures to show how the measures vary across months, which categories drive them, and how their composition evolves over time. Figure 2 plots each monthly venture capital allocation in (R_t, S_t) space. The horizontal axis measures within-direction concentration, and the vertical axis measures cross-direction relatedness. Moving northeast raises total relatedness, $\alpha_t = R_t + S_t$, because both components increase. Composition changes in a different direction: holding α_t fixed, moving toward the vertical axis raises s_t because a larger fraction of relatedness comes from funding across distinct directions. The line $R_t = S_t$ marks the balanced case $s_t = 0.5$; observations below it are dominated by within-direction concentration, while observations above it are dominated by cross-direction relatedness.

The four labeled months make the distinction concrete. January 2001 and September 2010 lie near the same constant-α_t line, so their total relatedness is similar. Their composition is not: $s_t = 0.390$ in January 2001, when within-direction concentration dominates, and $s_t = 0.500$ in September 2010, when the two components are balanced. November 2019 has the highest share among the labeled months, $s_t = 0.512$, even though both R_t and S_t are relatively low. December 2024 lies farther northeast, indicating substantially higher total relatedness, but its share is only 0.431 because within-direction concentration is larger than cross-direction relatedness. A high cross-direction share therefore need not coincide with high total relatedness.

Figure 3 shows which category cells generate these four points. The contributions are normalized by α_t, so the within-direction bars sum to R_t/α_t and the cross-direction bars sum to S_t/α_t. January 2001 is dominated by its diagonal cells: Internet and Software alone account for 24.2 and 11.9 percent of total relatedness. September 2010 is more balanced because several health-related category pairs contribute through cross-direction relatedness; Biotechnology–Health Care alone accounts for 10.0 percent. In November 2019, cross-direction relatedness is broad rather than concentrated. The displayed pairs are individually small, while the omitted pairs collectively account for 40.0 percent of total relatedness. December 2024 illustrates a different configuration. AI appears in both channels, but its within-direction cell accounts for 21.1 percent of total relatedness, compared with 5.5 percent for the largest AI cross-direction pair. The decline in s_t therefore reflects the relative strength of within-direction AI concentration, not the absence of AI-related connections across categories.

Figure 4 traces the same distinction over time. Panel (a) plots within-direction concentration R_t and cross-direction relatedness S_t; Panel (b) plots the cross-direction share s_t. The share begins at 0.390, rises toward the balanced value of 0.5, and remains near that level through much of the 2010s. It reaches 0.512 in November 2019 before falling to 0.431 by December 2024. Panel (a) explains the decline: both R_t and S_t rise near the end of the sample, but R_t rises faster. Total relatedness therefore increases while its composition shifts toward within-direction concentration. This divergence clarifies why s_t is not simply a rescaled version of S_t.

4 Results

4.1 Venture composition and public-market comovement

Baseline regressions. Implications 1 and 2 yield a joint prediction: venture composition should forecast both average pairwise correlation and the strength of the leading component, whereas total relatedness should have little predictive content once composition is held fixed. We test this prediction using the monthly regression

$$y_{t+h} = c + \beta_s\, s_t + \beta_\ell\, \alpha_t + X_t'\delta + \varepsilon_{t+h}, \tag{11}$$

where y_{t+h} is one of the two aggregate outcomes defined above, transformed using the Fisher mapping to place the bounded outcomes on a more convenient scale, and measured over the 12-month return window ending at $t + h$.[7] The baseline horizon is $h = 12$, so venture conditions at month t predict the public-market correlation structure over the following year. Both s_t and α_t are standardized. Thus, β_s measures the effect of a one-standard-deviation shift toward cross-direction relatedness, holding total relatedness fixed, while β_ℓ measures the effect of a one-standard-deviation change in total relatedness, holding composition fixed.

[7] The Fisher transformation is strictly increasing, so it preserves the signs of the model's comparative statics while rescaling their magnitudes; see Appendix B. Table A5 reports economically similar results using untransformed outcomes.

The baseline control vector X_t includes policy and market uncertainty, $\ln(EPU)$ and $\ln(VIX)$, together with financing conditions, $\ln(IPO+1)$ and $\ln(VC\ flow+1)$. Expanded specifications add investor sentiment and alternative measures of innovation or product-market change, including TNIC breadth, product-market fluidity, patent value, and BNBC News Attention (Tech Buzz). These controls address the possibility that both venture composition and future comovement respond to broader uncertainty, financing, sentiment, or innovation cycles. We report Newey–West standard errors with 12 monthly lags to account for serial correlation generated by overlapping return windows and persistent aggregate conditions.

Under the broadly same-signed exposure condition in Section 2, the model predicts $\beta_s > 0$ for both outcomes. A higher cross-direction share increases the common component of innovation news, raising the correlation of a typical firm pair; when the leading component is market-aligned, the same force raises the top-eigenvalue share. By contrast, the model predicts $\beta_\ell \approx 0$ because total relatedness does not enter the return-correlation equation (8). The combination of a positive share coefficient and a near-zero total-relatedness coefficient is therefore the sharper test of the composition mechanism.

Aggregate comovement results. Tables 2 and 3 report the main estimates for the two aggregate outcomes. Table 2 first examines the Fisher-transformed average pairwise correlation at $t+12$. The coefficient on the cross-direction share is positive and statistically significant in every specification, ranging from 0.046 to 0.075. With uncertainty and funding controls in column (1), a one-standard-deviation increase in the share predicts a 0.047 increase in the transformed outcome. The estimate rises to 0.075 after adding sentiment and TNIC breadth in column (2) and remains similar at 0.068 when patent value replaces TNIC breadth in column (3). Column (4) yields an estimate of 0.046 in the smaller BNBC news-attention sample, nearly identical to the baseline estimate despite the change in sample. Relative to the outcome standard deviation of 0.120 reported in Table 1, these coefficients represent approximately 38 to 63 percent of a standard deviation. The relation is therefore economically large as well as stable across the main control sets.

Total relatedness behaves differently. Its coefficients range from -0.020 to 0.005 and are statistically insignificant in all four columns. This share–total-relatedness contrast is the model's central placebo: increasing total relatedness does not predict higher public-market comovement once its composition is held fixed. What matters is whether that relatedness is generated by cross-direction relatedness rather than within-direction concentration. The results thus support Implication 1, which links a higher cross-direction share to stronger future average pairwise correlation under broadly same-signed innovation exposures.

Table 3 applies the same specifications to the Fisher-transformed top-eigenvalue share of the implied correlation matrix. The coefficient pattern closely mirrors the average-pairwise-correlation results: the estimates are 0.045, 0.071, 0.065, and 0.043 across columns (1)–(4), respectively, and each is statistically significant. These magnitudes correspond to approximately 36 to 60 percent of the outcome standard deviation of 0.119. The coefficient on total relatedness is again small and statistically insignificant, with estimates between -0.023 and 0.003.

The parallel results across the two outcomes sharpen the interpretation. A higher cross-direction share predicts not only greater average pairwise correlation, but also a return-correlation structure that places more weight on its leading component. That leading component is market-aligned in the data.[8] The second result is therefore the empirical counterpart to Implication 2: greater commonality in innovation news shifts a larger share of standardized return variation into the market direction. The close movement of the coefficients across columns also suggests that the two outcomes are responding to the same underlying composition signal rather than to separate features of the control variables.

Figure 5 extends this comparison across the full specification grid. For both outcomes, the share coefficient remains positive and its 95-percent confidence interval remains above zero as the controls are expanded to include uncertainty, funding activity, sentiment, TNIC breadth, product-market fluidity, patent value, and technology-news attention, as well as in specifications that exclude VC flow. This stability makes it less likely that the estimates reflect a general funding boom, elevated innovation activity, or broad market conditions that simultaneously affect venture capital and public-market comovement.

Taken together, the tables and Figure 5 support both aggregate predictions of the model. The same shift toward cross-direction relatedness forecasts higher average pairwise correlation and a larger market-aligned leading component, while total relatedness predicts neither outcome. The joint pattern is consistent with the proposed mechanism: when venture experimentation spans connected innovation directions, subsequent innovation news contains a stronger common component in public equity returns.

The horizon profile. Figure 6 traces the full-control estimate across return windows ending at t, $t + 6$, $t + 12$, $t + 18$, and $t + 24$. The model does not impose a particular horizon profile: the timing of the response can depend on how quickly innovations become relevant to public firms and are incorporated into prices. The figure therefore provides

[8] Using the leading eigenvector of the S&P 500 implied correlation matrix as portfolio weights, the resulting leading-component portfolio's return correlates 0.96 with the equal-weighted market and 0.92 with the value-weighted market over the sample. The leading component is therefore close to the equal-weighted market direction maintained in the model.

evidence about when the predictive relation is strongest. For average pairwise correlation, the share coefficient rises from 0.045 for the contemporaneous window to 0.070 at $t + 6$ and 0.075 at $t + 12$, then declines at $t + 18$ and $t + 24$. The top-eigenvalue share displays nearly the same hump-shaped pattern, increasing from approximately 0.042 at t to 0.066 at $t + 6$ and 0.071 at $t + 12$ before receding. Point estimates remain positive at every horizon, although the confidence intervals widen and include zero at $t + 24$. Thus, the relation is not solely contemporaneous: venture composition is most informative about the public-market correlation structure measured over the subsequent one-year window, consistent with innovation news reaching public firms with a lag.

Composition versus total relatedness. Figure 7 provides a partial-regression view of the distinction between composition and total relatedness. Panel (a) plots residualized average pairwise correlation against the residualized cross-direction share after removing the variation explained by total relatedness and the baseline controls. The standardized residual slope is 0.329, with a Newey–West t-statistic of 3.04. Panel (b) reverses the exercise: after removing variation associated with the share and the same controls, the residual slope on total relatedness is -0.097, with a t-statistic of -0.85. The figure therefore visualizes the same contrast as the baseline regressions. Conditional on total relatedness, its composition is strongly associated with subsequent comovement; conditional on composition, total relatedness has no systematic relation with the outcome. This evidence reinforces the model's share–total-relatedness distinction and shows that the result is visible in the underlying variation rather than being driven by a particular table specification.

Appendix Table A1 goes one step further by entering within-direction concentration and cross-direction relatedness separately. In the full-control specification for average pairwise correlation, the coefficient on within-direction concentration, R_t, is -0.162, while the coefficient on cross-direction relatedness, S_t, is 0.109; both are statistically significant at the 1-percent level. The top-eigenvalue results are similar: the corresponding coefficients are -0.155 and 0.102, significant at the 1- and 5-percent levels, respectively. The null result for total relatedness therefore masks opposing predictive relations between its two components. Within-direction concentration predicts weaker subsequent comovement, whereas cross-direction relatedness predicts stronger comovement and a larger leading component. This sign pattern maps directly into the theory: cross-direction experimentation makes innovation news more common across public firms, while within-direction concentration leaves more of its effect localized. Because $\alpha_t = R_t + S_t$ combines these opposing sources of relatedness, its aggregate predictive coefficient is small once composition is accounted for.

4.2 Robustness and additional evidence

The baseline results could reflect broader technology-funding patterns or particular measurement and sample choices rather than the composition of total relatedness itself. We address these concerns using technology-area funding controls, alternative measure constructions, outcome and inference checks, episode leave-outs, and placebo tests that break the observed category-similarity structure.

Table 4 summarizes the results. Unless noted otherwise, each exercise uses the full-control specification with TNIC breadth and includes standardized total relatedness. The tests ask whether the share coefficient survives alternative controls, constructions, inference choices, and sample periods.

Technology-area funding controls. The cross-direction share could proxy for shifts in the broad areas receiving venture funding rather than for the organization of funding across related directions. Panel (a) of Table 4 addresses this concern by adding flows into nine technology areas: biotech, healthcare, financial services, energy, AI and data, software, consumer, hardware, and business services.

With monthly area flows, the share coefficients are 0.058 for average pairwise correlation and 0.055 for the top-eigenvalue share, both significant at the 5-percent level. These estimates are somewhat below the baseline values of 0.075 and 0.071 but remain economically meaningful after conditioning on the broad allocation of venture capital across technology areas.

Using trailing 12-month area flows produces larger coefficients of 0.166 and 0.161, both significant at the 1-percent level. Both rows lead to the same conclusion: the result is not explained by which broad technology areas receive venture capital.[9]

Alternative measurement choices. Panels (b)–(d) of Table 4 vary the construction of the composition measure and similarity matrix, the transformation of the outcomes, and the treatment of serial correlation.

Panel (b) replaces the bounded share with the standardized log relatedness ratio $w_t = \ln(S_t/R_t)$.[10] Because w_t is monotone in s_t, it preserves the ordering of venture-composition

[9] The trailing area controls are constructed from the same funding rounds and window as s_t and α_t. They therefore absorb persistent variation in the funded-area mix mechanically and leave the share coefficient identified from a narrower residual comparison. This tighter conditioning raises both the coefficient and its standard error relative to the monthly-flow specification. We view the monthly-flow row as the cleaner area-control exercise and the trailing-flow row as a stringent complementary check.

[10] The log relatedness ratio is the standard log-ratio representation of a two-part composition (Aitchison, 1982).

states while using a different functional form. The coefficients are 0.075 for average pairwise correlation and 0.071 for the top-eigenvalue share, effectively identical to the baseline estimates. The result therefore does not depend on expressing composition as a bounded share.

Panel (c) estimates category similarity using longer category co-assignment windows. With a 24-month window, the coefficients are 0.052 and 0.050 for the two outcomes, both significant at the 5-percent level. With a 36-month window, both estimates are 0.071 and significant at the 1-percent level. Thus, the result is preserved when category similarity is inferred from a longer history of category co-assignment across startups.

Panel (d) uses the outcomes in their original, non-Fisher-transformed units and varies the Newey–West lag length. The raw-outcome coefficients remain positive and significant at 0.065 for average pairwise correlation and 0.061 for the top-eigenvalue share. Extending the adjustment from 12 to 18 and 24 lags leaves the point estimates at 0.075 and 0.071, with both remaining significant at the 1-percent level. The conclusions therefore do not depend on the outcome transformation or the baseline inference choice.

Major market and innovation episode leave-outs. The predictive relation could be concentrated in a few unusual periods when both venture funding and public-market comovement changed sharply. Panel (e) of Table 4 therefore excludes, one at a time, formation months from 2000–2003, 2008–2009, 2020–2021, and 2022–2023, covering the dot-com cycle, the financial crisis, the COVID-19 period, and the recent AI funding wave.

For average pairwise correlation, the share coefficients are 0.137, 0.064, 0.075, and 0.076 across the four exclusions; for the top-eigenvalue share, they are 0.137, 0.061, 0.071, and 0.071. Every estimate is positive and significant at the 1-percent level. Excluding the financial crisis, COVID-19 period, or recent AI wave leaves the estimates close to the full-sample values. Excluding 2000–2003 makes the relation substantially stronger, indicating that the dot-com period attenuates rather than generates the result. The predictive pattern is therefore not attributable to any single major episode.

Matrix-label placebo. The final exercise tests whether the observed category-similarity structure is essential to the result. In each of 1,000 placebo draws, Appendix Table A8 randomly permutes the mapping between industry categories and the category-similarity matrix, then reconstructs the cross-direction share and re-estimates the predictive regressions. Monthly funding shares across categories remain unchanged. The placebo therefore preserves when and where venture capital is invested, including aggregate funding cycles and category concentration, but breaks the category co-assignment links that determine which innovation directions are connected.

The placebo coefficients are centered near zero. In the full-control specifications, the observed coefficients are 0.075 for average pairwise correlation and 0.071 for the top-eigenvalue share, whereas the 95th percentile of both placebo distributions is 0.015. No relabeling produces a coefficient as large as the observed estimate, yielding one-sided placebo p-values of 0.001 for both outcomes. Appendix Figure A1 shows how the observed share differs over time from the placebo median and 5th–95th percentile range. Appendix Figure A2 then displays the regression evidence directly: the observed coefficients lie far to the right of the placebo distributions for both outcomes.

This test isolates the role of connectedness in the measure. Funding the same categories at the same times has no comparable predictive content when those categories are assigned arbitrary positions in the category-similarity matrix. The signal therefore comes from the actual match between venture allocations and the observed category co-assignment structure, as required by the mechanism, rather than from generic funding concentration or common time-series variation.

4.3 The cross-sectional anatomy of the composition effect

The aggregate results establish that the cross-direction share predicts higher average pairwise correlation and a larger leading component in public equity returns. Those findings do not by themselves reveal how broadly the effect is distributed. A sufficiently strong response among a small cluster of firms could raise both aggregate outcomes even if most of the correlation matrix were unaffected. Implication 3 imposes a more demanding cross-sectional restriction: because each pair's sensitivity is the exposure product $a_i a_j$, the aggregate effect should be reflected in the signs and magnitudes of many pairwise responses, not only in a few extreme coefficients.

We estimate a separate predictive regression for each firm pair:

$$\rho_{ij,t+h} = c_{ij} + \beta_{ij}^{s}\, s_t + \beta_{ij}^{\ell}\, \alpha_t + X_t' \delta_{ij} + u_{ij,t+h}, \tag{12}$$

where $\rho_{ij,t+h}$ is the realized correlation between firms i and j over the return window ending at $t + h$, with $h = 12$ in the baseline analysis. The cross-direction share s_t and total relatedness α_t are standardized, and X_t contains either the baseline or full control set used in the aggregate regressions. The coefficient β_{ij}^{s} is the empirical counterpart of the model sensitivity $a_i a_j$, allowing every pair to respond differently to venture composition.

The sign restriction is informative. Pairs whose firms have same-signed innovation exposures should have $\beta_{ij}^{s} > 0$, whereas pairs containing a beneficiary and a displaced firm should have $\beta^s_{ij} < 0$. The theory therefore does not require every coefficient to be positive. It predicts that the distribution should be tilted toward positive values when exposures are broadly same-signed. Because the pair estimates share firms and time-series regressors and are noisy individually, we use their distribution and concentration as the relevant anatomy rather than treating pair-by-pair significance as independent evidence.

We then aggregate the pair coefficients to the firm level. Assigning half of each pair response to each of its members gives firm i's contribution

$$k_i = \frac{1}{2} \sum_{j \neq i} \beta^s_{ij}. \tag{13}$$

This construction is an exact allocation of the summed pairwise effect: $\sum_i k_i = \sum_{i<j} \beta^s_{ij}$. In the model, k_i corresponds to $\frac{1}{2} a_i \sum_{j \neq i} a_j$. Thus, a positive value means that firm i's net contribution across all of its pairs reinforces the aggregate comovement response. The firm-level distribution asks whether the aggregate is supported by many firms or carried by a small set of unusually exposed names.

Pair-level responses. Panel (a) of Figure 8 displays the full-control distribution across 156,889 firm pairs. The distribution is centered to the right of zero: 77.6 percent of pair coefficients are positive. Table 5 shows that the positive share is already 70.66 percent with baseline controls and rises to 77.61 percent with the full control set. The median coefficient increases from 0.04 to 0.07, while the mean increases from 0.04 to 0.08. The close agreement between the mean and median shows that the positive response is a feature of the center of the distribution rather than the product of a small right tail. The pair coefficients are expressed in pair-correlation units, whereas the main aggregate outcome is Fisher-transformed, so their magnitudes are not directly comparable; their common positive direction is the relevant cross-sectional restriction.

The negative coefficients remain economically informative rather than contradicting the mechanism. Approximately 22 percent of pairs respond negatively in the full-control specification, consistent with the model allowing some firms to benefit from a connected innovation stack while others are displaced. What matters for the aggregate prediction is that same-signed exposure products dominate, which is precisely what the positive shift in the distribution indicates. Moreover, the pattern becomes stronger after adding sentiment and TNIC breadth to the controls.

The concentration evidence rules out an explanation based on a handful of exceptional pairs. The 100 largest positive pair coefficients account for 0.57 percent of total positive coefficient mass under baseline controls and 0.58 percent under full controls. These 100 observations represent less than one-tenth of one percent of all estimated pairs, yet even the most extreme among them explain well under 1 percent of the positive response. The aggregate coefficient therefore cannot be attributed to a small collection of unusually exposed bilateral relationships.

Firm-level contributions. Panel (b) of Figure 8 shows an even stronger positive tilt after pair responses are aggregated within firms. Among 690 firms, 78.99 percent have positive contributions under baseline controls and 84.93 percent do so under full controls. The higher positive share at the firm level means that, although a firm may belong to some negatively responding pairs, its net contribution across all relationships is positive for the large majority of firms.

The magnitudes reinforce this conclusion. The median firm contribution rises from 10.10 in the baseline specification to 20.06 with full controls; the corresponding means are 9.70 and 17.78. Firm contributions are sums over many pair coefficients and are therefore not directly comparable in scale to an individual pair response. Within the firm distribution, however, the positive medians show that the typical firm contributes in the same direction as the aggregate effect. The median exceeds the mean in both specifications, providing little indication that a few very large positive firms are pulling an otherwise weak distribution upward.

Firm contributions are more concentrated than pair coefficients, as the model would naturally allow: a firm with a large exposure enters many pair products through the same a_i. Even so, the concentration is limited. The 100 largest positive firm contributions account for 37.73 percent of positive mass under baseline controls and 31.87 percent under full controls. Thus, in the full specification, the remaining firms account for more than two-thirds of the positive contribution mass. The decline in the top-100 share after adding controls further indicates that the broad firm-level pattern is not generated by a small set of dominant names.

Taken together, the results support all parts of Implication 3. The majority of pair sensitivities are positive, both the mean and median pair responses are positive, most firms make positive net contributions, and neither pair nor firm contribution mass is dominated by the largest observations. These facts add cross-sectional discipline to the aggregate findings. The same venture-composition signal raises average pairwise correlation, strengthens the market-aligned leading component, and appears throughout the pair and firm distributions in the form predicted by a common vector of innovation exposures. A narrow exposure story could generate a positive aggregate coefficient, but it would not naturally produce this combination of widespread positive signs and diffuse contribution mass. The evidence instead points to innovation news entering public returns through a broad common component while still allowing a minority of firms and pairs to respond in the opposite direction.

5 Conclusion

We show that the composition of venture experimentation forecasts whether innovation risk becomes systematic or remains firm-specific in public markets. When a larger share of total relatedness comes from funding across distinct but connected innovation directions, average pairwise correlation rises over the following year and more return variation is absorbed by the market-aligned leading component. Total relatedness does not carry the same signal. The response is broad across firm pairs and firms, survives the main controls and measurement alternatives, and disappears when the observed category co-assignment links among innovation directions are randomly broken.

For asset pricing, the results suggest that factor structure is partly formed before the relevant technologies are fully represented in public markets. Common factors need not arise only from contemporaneous macroeconomic shocks, trading behavior, or linkages among listed firms. They can also emerge as private experimentation builds a connected innovation stack whose value is exposed to common innovation news. This perspective makes the boundary between firm-specific and systematic risk endogenous to the organization of technological development and identifies private capital markets as an upstream source of information about future covariance.

For venture capital, the findings shift attention from funding quantity and sector allocation to portfolio architecture. Two periods with similar funding volumes and broad technology-area exposures can have different implications if one finances complementary layers of a connected innovation stack while the other concentrates capital within a narrow direction. This distinction also matters for diversification within venture portfolios: investments classified as different businesses may share substantial common exposure when they occupy connected layers of the same innovation stack. Measuring relatedness across innovation directions can therefore reveal portfolio dependence that conventional company or sector labels miss.

The policy and investment implications are not that cross-direction funding is inherently desirable or undesirable. Financing a connected innovation stack may create valuable complementarities and accelerate diffusion, while also making the resulting innovation risk harder to diversify in public markets. Conversely, concentrated experimentation may produce important breakthroughs whose valuation effects remain more firm-specific. For investors, policymakers, and institutions monitoring technological change, the practical lesson is to track how capital connects innovation directions alongside how much capital is deployed. Because our evidence is predictive rather than causal, using this signal for intervention requires further evidence on the welfare tradeoff between technological spillovers, concentration, and the systematic risk created as innovation stacks mature.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society, Series B*, 44(2):139–177.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4):1645–1680.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4):1593–1636.

Barberis, N., Shleifer, A., and Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, 75(2):283–317.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393.

Bresnahan, T. F. and Trajtenberg, M. (1995). General purpose technologies: "engines of growth"? *Journal of Econometrics*, 65(1):83–108.

Bybee, L., Kelly, B. T., Manela, A., and Xiu, D. (2024). Business news and business cycles. *Journal of Finance*, 79(5):3105–3147.

Campbell, J. Y., Lettau, M., Malkiel, B. G., and Xu, Y. (2001). Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. *Journal of Finance*, 56(1):1–43.

Driessen, J., Maenhout, P. J., and Vilkov, G. (2009). The price of correlation risk: Evidence from equity options. *Journal of Finance*, 64(3):1377–1406.

Engle, R. F. and Kelly, B. T. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, 30(2):212–228.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772.

Gârleanu, N., Kogan, L., and Panageas, S. (2012). Displacement risk and asset returns. *Journal of Financial Economics*, 105(3):491–510.

Herskovic, B. (2018). Networks in production: Asset pricing implications. *Journal of Finance*, 73(4):1785–1818.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, 119(2):249–283.

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

Hoberg, G., Phillips, G., and Prabhala, N. R. (2014). Product market threats, payouts, and financial flexibility. *Journal of Finance*, 69(1):293–324.

Hobijn, B. and Jovanovic, B. (2001). The information-technology revolution and the stock market: Evidence. *American Economic Review*, 91(5):1203–1220.

Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3):303–320.

Kogan, L. and Papanikolaou, D. (2014). Growth opportunities, technology shocks, and asset prices. *Journal of Finance*, 69(2):675–718.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *Quarterly Journal of Economics*, 132(2):665–712.

Kortum, S. and Lerner, J. (2000). Assessing the contribution of venture capital to innovation. *RAND Journal of Economics*, 31(4):674–692.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411.

Lerner, J. and Nanda, R. (2020). Venture capital's role in financing innovation: What we know and how much we still need to learn. *Journal of Economic Perspectives*, 34(3):237–261.

Morck, R., Yeung, B., and Yu, W. (2000). The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial Economics*, 58(1–2):215–260.

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

Pástor, L. and Veronesi, P. (2009). Technological revolutions and stock prices. *American Economic Review*, 99(4):1451–1483.

Pindyck, R. S. and Rotemberg, J. J. (1993). The comovement of stock prices. *Quarterly Journal of Economics*, 108(4):1073–1104.

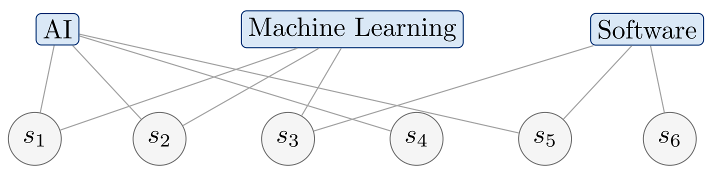
Pollet, J. M. and Wilson, M. (2010). Average correlation and stock market returns. *Journal of Financial Economics*, 96(3):364–380.

Samila, S. and Sorenson, O. (2011). Venture capital, entrepreneurship, and economic growth. *Review of Economics and Statistics*, 93(1):338–349.

Veldkamp, L. L. (2006). Information markets and the comovement of asset prices. *Review of Economic Studies*, 73(3):823–845.

Figure 1: Building the measure: a three-category example. This figure shows how startup industry-category assignments are transformed into measures of venture capital relatedness. Panel (a) illustrates a toy economy with six startups and three innovation directions: AI, Machine Learning, and Software. Category co-assignment reveals which directions are empirically close. Panel (b) reports the resulting similarity matrix, where diagonal entries are normalized to one and off-diagonal entries capture similarity across distinct directions. Panel (c) combines this matrix with a venture capital allocation vector to form the contribution matrix $C_{k\ell,t} = p_{kt}p_{\ell t}W_{k\ell}$. Diagonal cells sum to within-direction concentration R_t, while off-diagonal cells sum to cross-direction relatedness S_t.

Panel (a): Startup-category assignments



Panel (b): Similarity matrix

	AI	ML	Soft.
AI	1	0.40	0.17
ML	0.40	1	0.20
Soft.	0.17	0.20	1

Own-category similarity, normalized to one.
Strong revealed category similarity.
Weaker revealed category similarity.

Panel (c): Venture capital allocation contribution matrix

$$p_t = (p_{\mathrm{AI}}, p_{\mathrm{ML}}, p_{\mathrm{Soft.}}) = (0.40, 0.35, 0.25)$$

	AI	ML	Soft.
AI	0.1600	0.0560	0.0170
ML	0.0560	0.1225	0.0175
Soft.	0.0170	0.0175	0.0625

Diagonal cells: R_t, within-direction concentration.
Large off-diagonal cells: stronger contribution to S_t.
Small off-diagonal cells: weaker contribution to S_t.

Figure 2: Monthly venture capital allocation states in (R_t, S_t) space. Each point is a trailing 12-month endpoint, and color records calendar year. The axes use identical scales. The solid line marks $R_t = S_t$. The dashed line is centered on the January 2001–September 2010 pair and reports their nearly equal total relatedness, $\alpha_t \simeq 0.04$. The four labeled months are decomposed in Figure 3. December 2024 is a descriptive extension beyond the predictive formation sample.

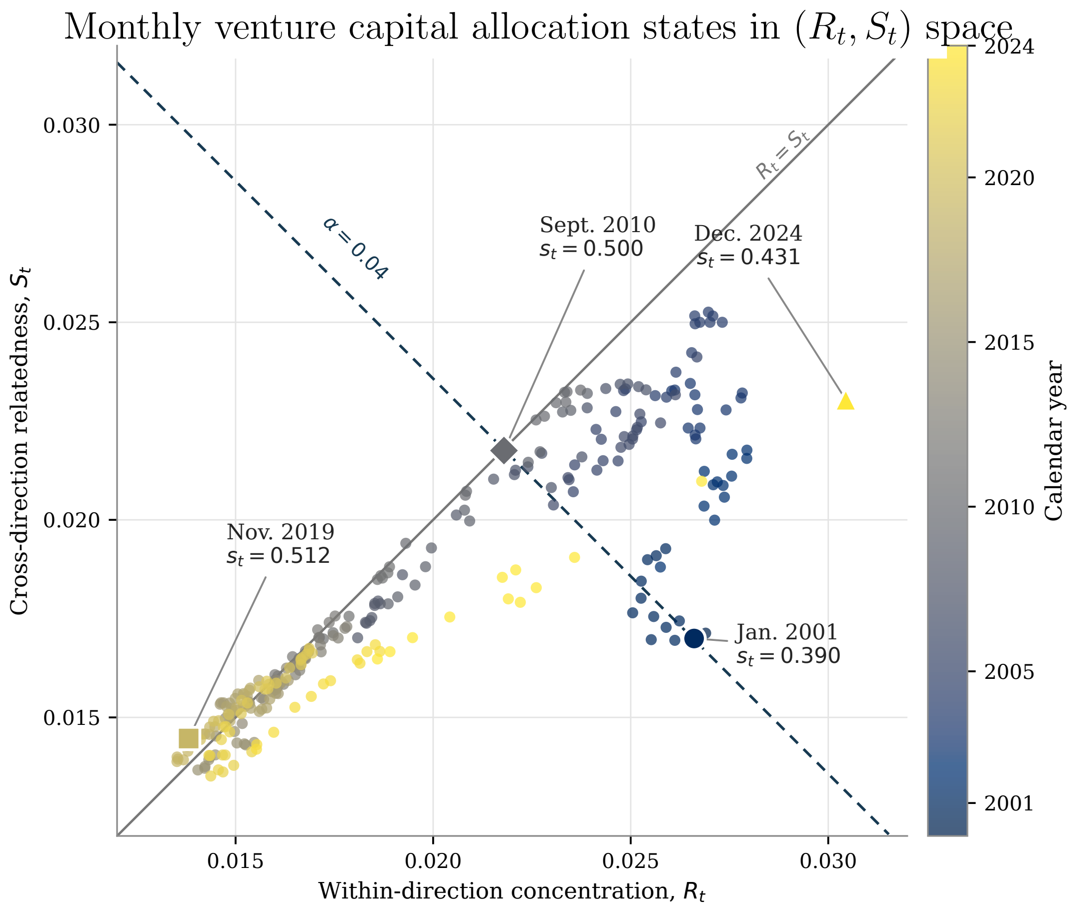


Figure 3: Leading contributors to within-direction concentration and cross-direction relatedness. Diagonal bars report p_{kt}^2/α_t and sum to R_t/α_t within each month. Cross-direction bars report $2p_{kt}p_{\ell t}W_{k\ell,t}/\alpha_t$ for unordered category pairs and sum to S_t/α_t. "All remaining" includes every omitted cell in the corresponding channel. Contributions use a common scale across dates and channels. December 2024 is a descriptive extension beyond the predictive formation sample.

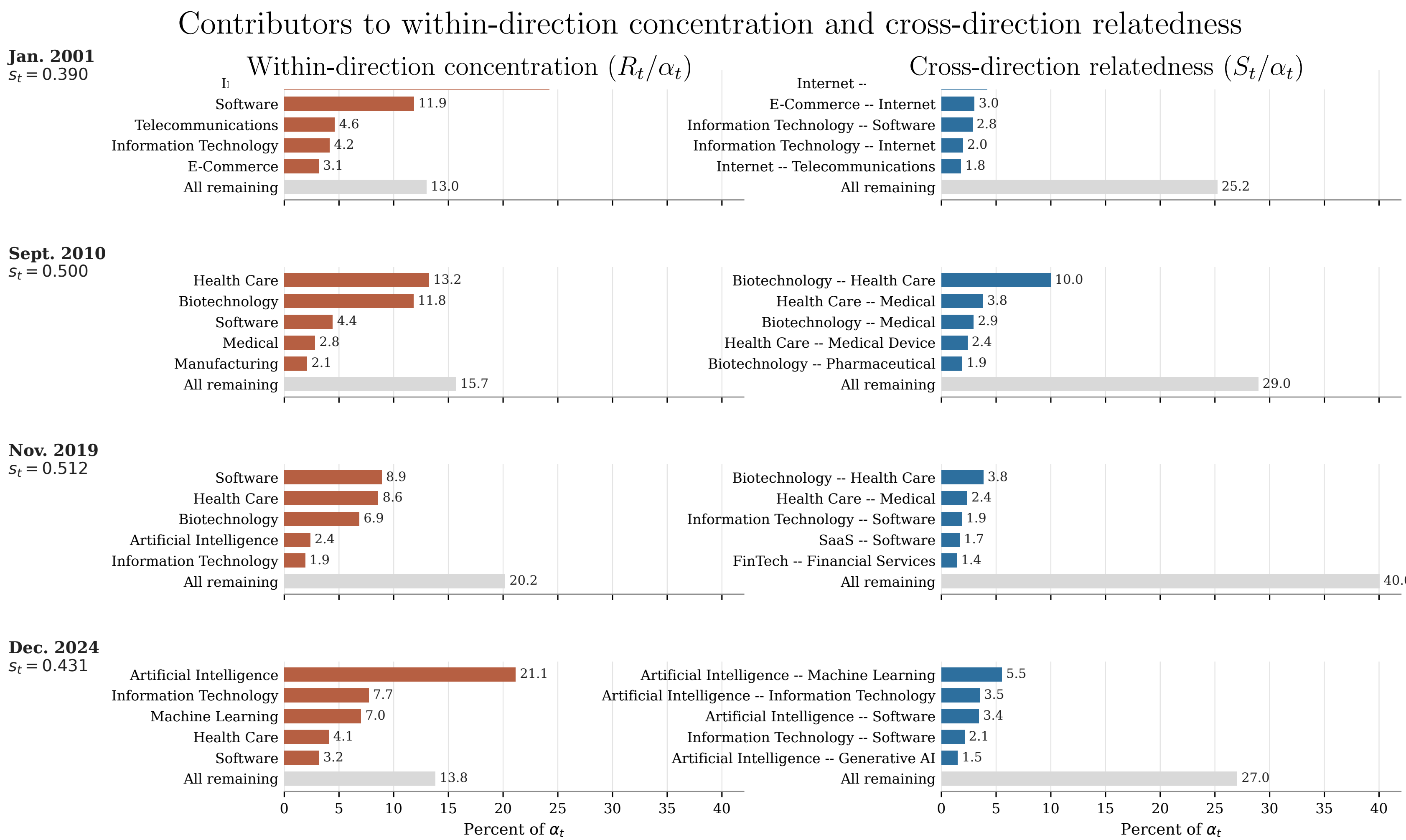


Figure 4: Components and cross-direction share of total relatedness. Panel (a) plots within-direction concentration R_t and cross-direction relatedness S_t in their original units. Panel (b) plots the cross-direction share $s_t = S_t/(R_t + S_t)$. The horizontal reference at $s_t = 0.5$ means $R_t = S_t$; it does not mean that funding is equally allocated across categories. Vertical guides mark the four months highlighted in Figures 2 and 3. All series use trailing 12-month windows. December 2024 is a descriptive extension beyond the predictive formation sample.

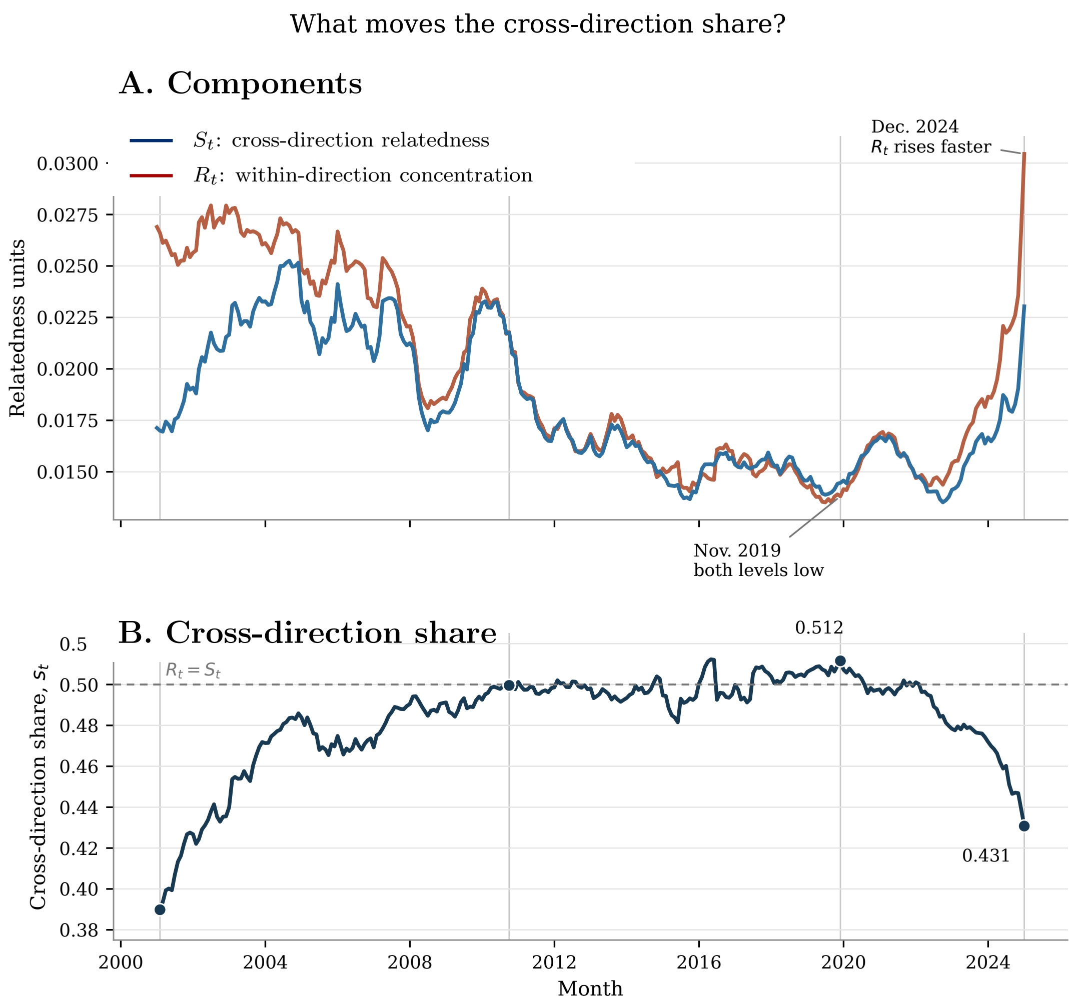
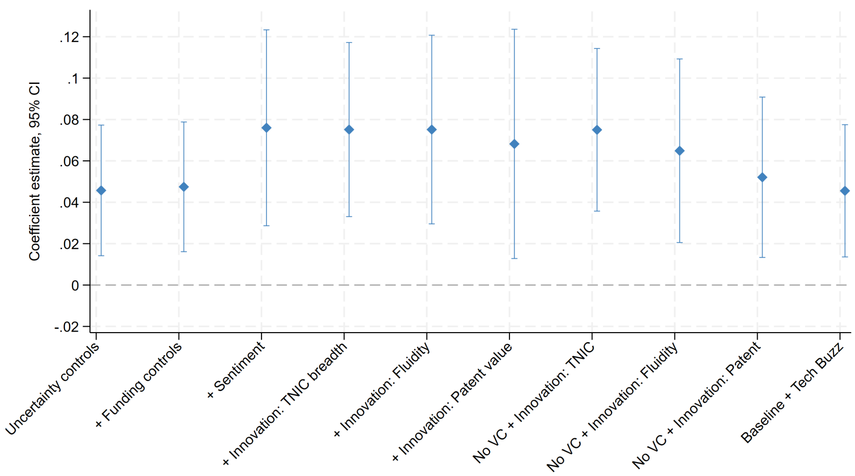


Figure 5: The share coefficient across specifications. This figure reports the stability of the main estimate across the full predictive specification grid. Each point is the Newey–West coefficient on the standardized cross-direction share s_t from a monthly regression of the 12-month-ahead outcome on the share and standardized total relatedness, with regressors and controls measured at time t. Panel (a) uses the Fisher-transformed average pairwise return correlation at $t + 12$; Panel (b) uses the Fisher-transformed top-eigenvalue share of the implied correlation matrix at $t + 12$. Whiskers denote 95 percent confidence intervals based on Newey–West standard errors with 12 lags. The specifications progressively add uncertainty controls, funding controls, sentiment, TNIC breadth, product-market fluidity, patent value, the BNBC News Attention (Tech Buzz) control, and variants excluding VC flow.

Panel (a): Average pairwise correlation



Panel (b): Top-eigenvalue share

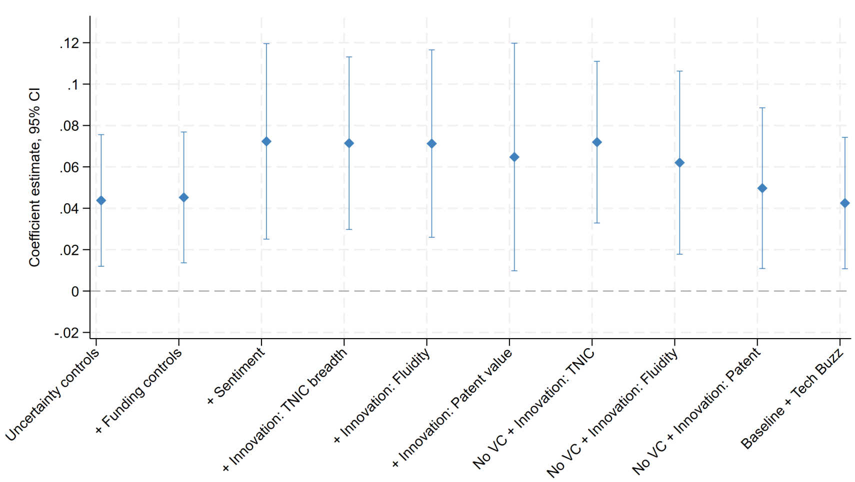
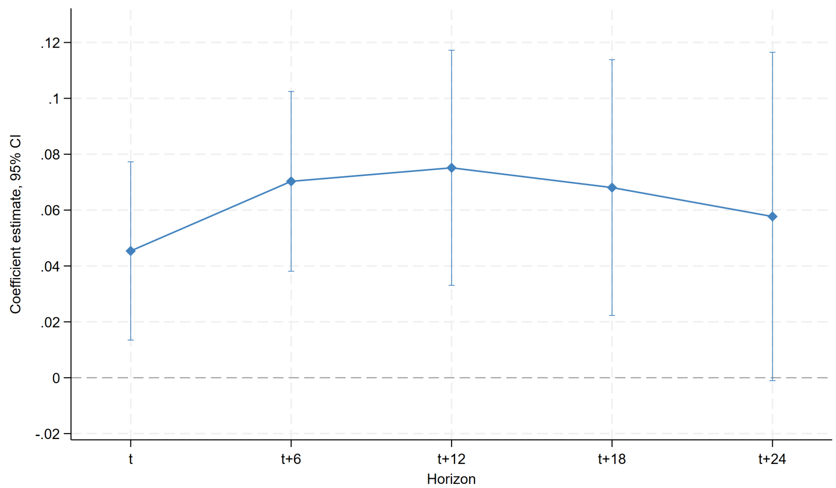


Figure 6: The horizon profile. This figure traces the coefficient on the standardized cross-direction share s_t across public-market outcomes measured over windows ending at t, $t+6$, $t+12$, $t+18$, and $t+24$. Each point comes from the specification with uncertainty, funding, sentiment, and TNIC-breadth controls, with standardized total relatedness included as a regressor. Panel (a) uses the Fisher-transformed average pairwise return correlation; Panel (b) uses the Fisher-transformed top-eigenvalue share of the implied correlation matrix. Whiskers denote 95 percent confidence intervals based on Newey–West standard errors with 12 lags.

Panel (a): Average pairwise correlation



Panel (b): Top-eigenvalue share

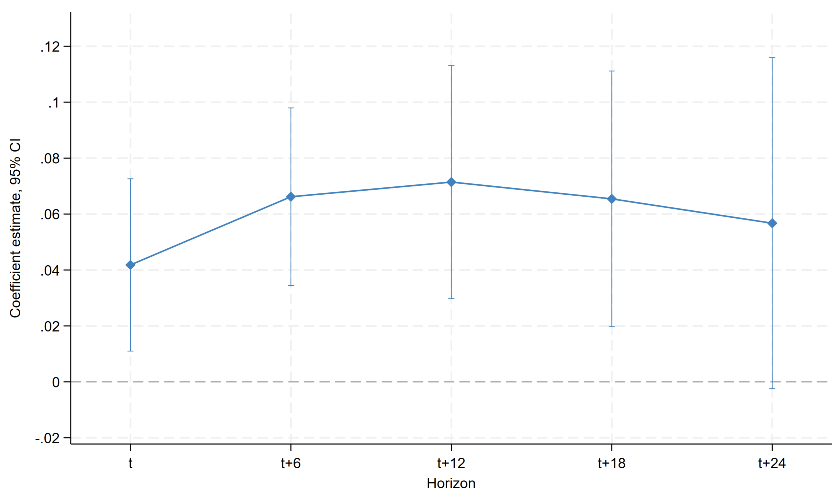
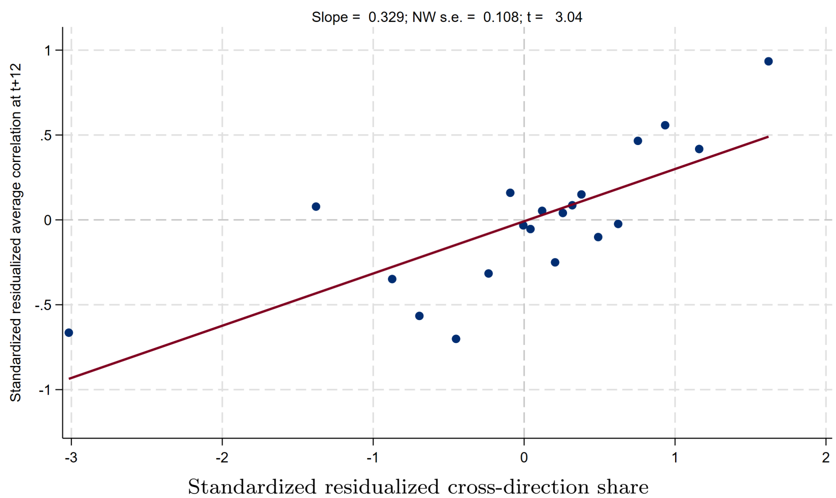


Figure 7: Decomposing relatedness: cross-direction share versus total relatedness. This figure separates the cross-direction share from total relatedness. Panel (a) plots standardized residualized average pairwise return correlation measured at $t + 12$ against the standardized residualized cross-direction share, holding total relatedness and baseline controls fixed. Panel (b) plots the same outcome against standardized residualized total relatedness, holding the cross-direction share and the same controls fixed. Both axes use standardized residuals from the corresponding Frisch–Waugh–Lovell residualization. The fitted lines report slopes from residual-on-residual regressions, and standard errors are Newey–West with 12 monthly lags.

Panel (a): Average pairwise correlation against the cross-direction share



Panel (b): Average pairwise correlation against total relatedness

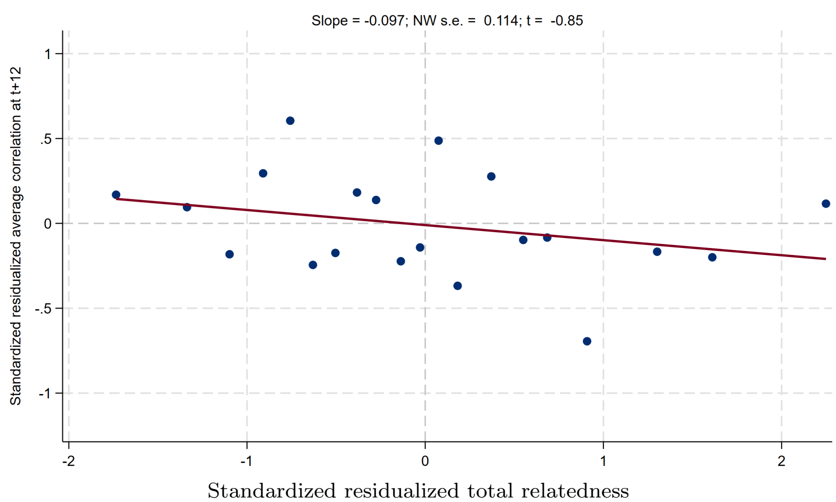
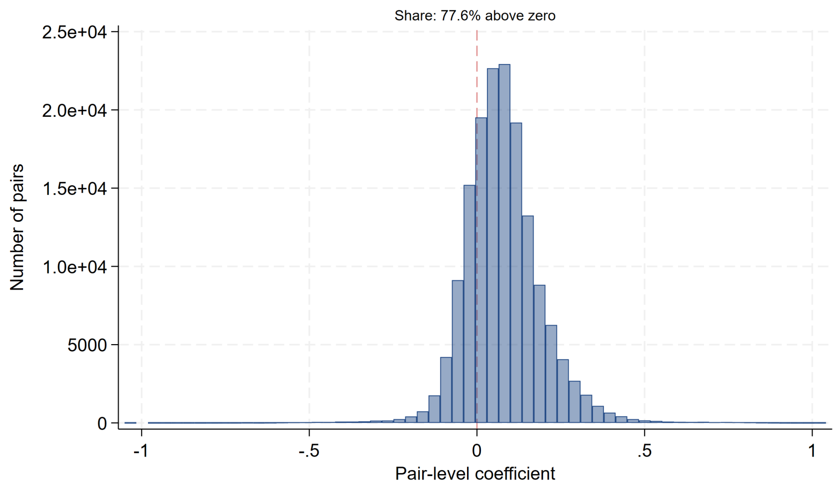


Figure 8: Pair-level and firm-level anatomy. This figure reports the cross-sectional anatomy of the composition effect, from pair-by-pair regressions of pair correlations at $t + 12$ on the standardized cross-direction share s_t and standardized total relatedness α_t, with the full control set (uncertainty, funding, sentiment, and TNIC breadth). Panel (a) shows the distribution of pair-level share coefficients across firm pairs; 77.6 percent are positive. Panel (b) shows the distribution of firm-level contributions, assigning one half of each pair coefficient to each firm; 84.9 percent are positive.

Panel (a): Distribution of pair-level share coefficients



Panel (b): Distribution of firm-level contributions

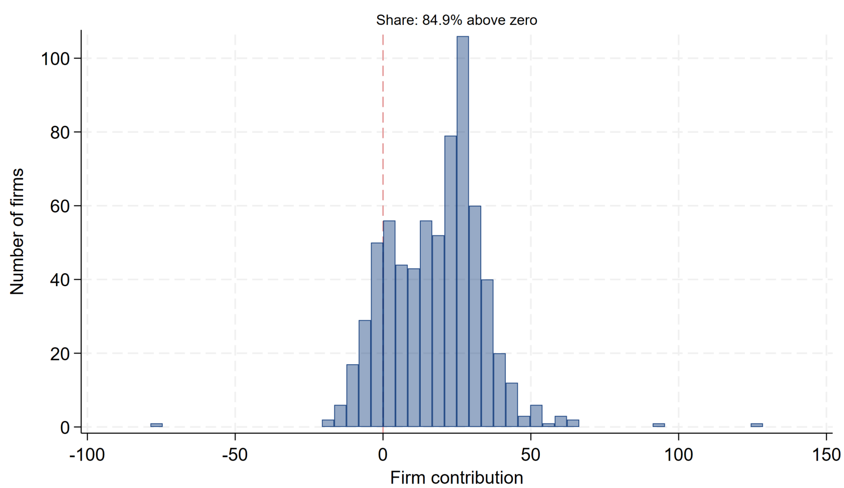


Table 1: Summary statistics. This table presents summary statistics for the monthly variables used in the baseline analysis. The baseline sample contains 277 monthly observations, with smaller samples for variables available over shorter windows. Public-market outcomes are measured at $t+12$ and reported both in levels and in the Fisher-transformed units used in the baseline regressions.

	N	Mean	SD	Median	Min	Max
Venture composition variables						
Cross-direction share (s_t)	277	0.484	0.024	0.493	0.389	0.512
Within-direction concentration (R_t)	277	0.019	0.005	0.017	0.014	0.028
Cross-direction relatedness (S_t)	277	0.018	0.003	0.017	0.014	0.025
Total relatedness (α_t)	277	0.037	0.008	0.035	0.027	0.052
Log relatedness ratio ($w_t = \ln(S_t/R_t)$)	277	-0.063	0.098	-0.029	-0.451	0.049
Outcome variables						
Average pairwise correlation	277	0.307	0.105	0.291	0.101	0.545
Average pairwise correlation, Fisher transformed	277	0.322	0.120	0.299	0.102	0.611
Top-eigenvalue share	277	0.326	0.102	0.310	0.124	0.557
Top-eigenvalue share, Fisher transformed	277	0.343	0.119	0.321	0.124	0.628
Control variables						
ln(EPU)	277	4.585	0.449	4.540	3.618	6.221
ln(VIX)	277	2.924	0.355	2.877	2.315	4.138
$\ln(\text{IPO}+1)$	277	2.763	0.773	2.833	0.000	5.024
$\ln(\text{VC flow}+1)$	277	8.286	0.898	8.132	6.565	10.367
Investor sentiment	277	0.124	0.678	-0.030	-0.968	3.019
TNIC breadth	277	0.734	0.042	0.737	0.637	0.795
Product-market fluidity	277	6.552	1.323	6.740	3.267	9.459
ln(patent value)	277	11.361	0.464	11.222	10.541	12.521
BNBC news attention (Tech Buzz)	199	0.042	0.006	0.041	0.029	0.056

Table 2: Venture composition and average pairwise correlation. This table presents the results of monthly time-series regressions where the dependent variable is the Fisher-transformed average pairwise return correlation measured at $t+12$. Regressors and controls are measured at time t. The cross-direction share $s_t = S_t/(S_t+R_t)$ and total relatedness $\alpha_t = S_t+R_t$ are standardized. Column (1) includes uncertainty and funding controls; columns (2)–(4) add additional controls. Uncertainty controls are $\ln(EPU)$ and $\ln(VIX)$; funding controls are $\ln(IPO+1)$ and $\ln(VC\ flow+1)$; the innovation control is TNIC breadth, patent value, or BNBC news attention (Tech Buzz), one at a time. Standard errors are Newey–West with 12 monthly lags and are reported in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

	(1)	(2)	(3)	(4)
Cross-direction share (s_t)	0.047***	0.075***	0.068**	0.046***
	(0.016)	(0.021)	(0.028)	(0.016)
Total relatedness (α_t)	−0.020	−0.011	−0.016	0.005
	(0.022)	(0.021)	(0.023)	(0.021)
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment	No	Yes	Yes	No
Innovation control	No	TNIC	Patent	Tech Buzz
Observations	277	277	277	199

Table 3: Venture composition and the top-eigenvalue share of the implied correlation matrix. This table presents the results of monthly time-series regressions where the dependent variable is the Fisher-transformed top-eigenvalue share of the implied correlation matrix measured at $t+12$. Regressors and controls are measured at time t. The cross-direction share $s_t = S_t/(S_t + R_t)$ and total relatedness $\alpha_t = S_t + R_t$ are standardized. Column (1) includes uncertainty and funding controls; columns (2)–(4) add additional controls. Uncertainty controls are $\ln(EPU)$ and $\ln(VIX)$; funding controls are $\ln(IPO + 1)$ and $\ln(VC\ flow + 1)$; the innovation control is TNIC breadth, patent value, or BNBC news attention (Tech Buzz), one at a time. Standard errors are Newey–West with 12 monthly lags and are reported in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

	(1)	(2)	(3)	(4)
Cross-direction share (s_t)	0.045***	0.071***	0.065**	0.043***
	(0.016)	(0.021)	(0.028)	(0.016)
Total relatedness (α_t)	−0.023	−0.013	−0.019	0.003
	(0.022)	(0.021)	(0.023)	(0.021)
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment	No	Yes	Yes	No
Innovation control	No	TNIC	Patent	Tech Buzz
Observations	277	277	277	199

Table 4: Robustness summary. This table summarizes selected robustness checks for the coefficient on the standardized cross-direction share. The two outcome columns report estimates for average pairwise correlation and the top-eigenvalue share of the implied correlation matrix, measured at $t + 12$ and Fisher-transformed. All specifications include standardized total relatedness. The reference row reports the main full-control specification with uncertainty, funding, sentiment, and TNIC-breadth controls presented in Tables 2 and 3. The Newey–West row in Panel (d) reports the full-control estimates under alternative lag choices; point estimates are unchanged relative to the baseline, and stars correspond to the weakest significance level across the alternative lag specifications. Detailed estimates are reported in the appendix tables listed in the final column.

Robustness check	Average pairwise correlation	Top-eigenvalue share	Appendix table
Reference estimate			
Main full-control estimate	0.075***	0.071***	Tables 2–3
Panel (a): Technology-area controls			
Monthly technology-area flows	0.058**	0.055**	Table A2, Panel (a)
Trailing 12-month technology-area flows	0.166***	0.161***	Table A2, Panel (b)
Panel (b): Alternative composition measure			
Log relatedness ratio	0.075***	0.071***	Table A3
Panel (c): Alternative similarity windows			
24-month similarity window	0.052**	0.050**	Table A4
36-month similarity window	0.071***	0.071***	Table A4
Panel (d): Outcome transformation and inference			
Non-Fisher outcomes	0.065***	0.061***	Table A5
Alternative Newey–West lags	0.075***	0.071***	Table A6
Panel (e): Episode leave-outs			
Exclude 2000–2003	0.137***	0.137***	Table A7
Exclude 2008–2009	0.064***	0.061***	Table A7
Exclude 2020–2021	0.075***	0.071***	Table A7
Exclude 2022–2023	0.076***	0.071***	Table A7

Table 5: Pair-level and firm-level anatomy. This table summarizes the cross-sectional anatomy of the composition effect from pair-by-pair regressions of pair correlations at $t+12$ on the standardized cross-direction share s_t and standardized total relatedness α_t. Baseline controls are uncertainty and funding; full controls add sentiment and TNIC breadth. Firm contributions assign one half of each pair coefficient to each of the two firms in the pair. The top-100 rows report the share of total positive contribution mass accounted for by the 100 largest positive pair or firm contributions.

	Baseline controls	Full controls
Firm pairs		
Number of pairs	156,889	156,889
Positive share coefficients (%)	70.66	77.61
Median share coefficient	0.04	0.07
Average share coefficient	0.04	0.08
Top-100 pairs' share of positive contribution mass (%)	0.57	0.58
Firms		
Number of firms	690	690
Positive contributions (%)	78.99	84.93
Median contribution	10.10	20.06
Average contribution	9.70	17.78
Top-100 firms' share of positive contribution mass (%)	37.73	31.87

A Derivations and Proofs

The central equation and normalization. The model in equations (6)–(7) is

$$x_{i,t}^{(h)} = \sqrt{1-a_i^2}\, e_{i,t}^{(h)} + a_i\, g_{i,t}^{(h)}, \qquad g_{i,t}^{(h)} = \sqrt{s_t}\, f_t^{(h)} + \sqrt{1-s_t}\, u_{i,t}^{(h)},$$

with $e_{i,t}^{(h)}$, $f_t^{(h)}$, and $u_{i,t}^{(h)}$ mutually orthogonal and of unit variance, and $\mathrm{Cov}(u_{i,t}^{(h)}, u_{j,t}^{(h)}) = 0$ for $i \neq j$. Innovation news is itself unit-variance,

$$\mathrm{Var}(g_{i,t}^{(h)}) = s_t\ \mathrm{Var}(f_t^{(h)}) + (1-s_t)\ \mathrm{Var}(u_{i,t}^{(h)}) = 1,$$

and the two loadings in $x_{i,t}^{(h)}$ square to one, so

$$\mathrm{Var}(x_{i,t}^{(h)}) = (1-a_i^2) + a_i^2 = 1.$$

For $i \neq j$, the firm-specific shocks are uncorrelated and drop out of the cross-firm covariance, leaving the common shock and the non-innovation comovement:

$$\mathrm{Cov}(x_{i,t}^{(h)}, x_{j,t}^{(h)}) = \sqrt{1-a_i^2}\sqrt{1-a_j^2}\ \mathrm{Cov}(e_{i,t}^{(h)}, e_{j,t}^{(h)}) + a_i a_j\, s_t \equiv \rho_{ij,t}^{0,(h)} + s_t a_i a_j,$$

where $\rho_{ij,t}^{0,(h)}$ collects the non-innovation contribution to the correlation of x: the correlation of the underlying shocks $e_{i,t}^{(h)}$ and $e_{j,t}^{(h)}$, multiplied by $\sqrt{1-a_i^2}\sqrt{1-a_j^2}$. Because each return has unit variance, this covariance is also the correlation, which proves equation (8).

Derivation for Implication 1. Averaging (8) over ordered firm pairs gives

$$\bar{\rho}_t^{(h)} = \frac{1}{N(N-1)} \sum_{i \neq j} \left(\rho_{ij,t}^{0,(h)} + s_t a_i a_j \right) = \bar{\rho}_t^{0,(h)} + \frac{s_t}{N(N-1)} \sum_{i \neq j} a_i a_j,$$

where $\bar{\rho}_t^{0,(h)} \equiv \frac{1}{N(N-1)} \sum_{i \neq j} \rho_{ij,t}^{0,(h)}$ is fixed in the comparative static. Differentiating with respect to s_t gives

$$\frac{\partial \bar{\rho}_t^{(h)}}{\partial s_t} = \frac{1}{N(N-1)} \sum_{i \neq j} a_i a_j, \tag{14}$$

which is positive exactly when $\sum_{i \neq j} a_i a_j > 0$. A higher cross-direction share therefore raises average pairwise correlation when exposures are broadly same-signed, the claim in Implication 1.

Derivation for Implication 2. Let $\mathcal{C}_t^{(h)}$ be the predictive correlation matrix. Since the diagonal of a correlation matrix is fixed at one, the composition-sensitive part has zero diagonal and off-diagonal entries $a_i a_j$. Thus

$$\mathcal{C}_t^{(h)} = \mathcal{C}_t^{0,(h)} + s_t K, \qquad K \equiv aa' - \operatorname{diag}(a_i^2).$$

If the largest eigenvalue is simple, standard eigenvalue perturbation gives

$$\frac{\partial \lambda_1(\mathcal{C}_t^{(h)})}{\partial s_t} = v_1' K v_1 = (v_1' a)^2 - \sum_i v_{1i}^2 a_i^2.$$

The denominator in the top-eigenvalue share is N, the trace of any correlation matrix, so

$$\frac{\partial}{\partial s_t}\left(\frac{\lambda_1(\mathcal{C}_t^{(h)})}{N}\right) = \frac{1}{N}\left[(v_1' a)^2 - \sum_i v_{1i}^2 a_i^2\right]. \tag{15}$$

At the market direction $m = N^{-1/2}\mathbf{1}$,

$$m' K m = \frac{1}{N}\Big(\sum_i a_i\Big)^2 - \frac{1}{N}\sum_i a_i^2 = \frac{1}{N}\sum_{i \neq j} a_i a_j,$$

so

$$\frac{\partial}{\partial s_t}\left(\frac{\lambda_1(\mathcal{C}_t^{(h)})}{N}\right) = \frac{1}{N^2}\sum_{i \neq j} a_i a_j \tag{16}$$

when $v_1 = m$. If the average exposure product is positive, this derivative is positive at m; by continuity of $v' K v$, it stays positive for first eigenvectors sufficiently close to the market direction. This is the content of Implication 2: under broadly same-signed exposures the leading component's share rises and its eigenvector remains aligned with the market, rather than a new sectoral factor forming.

Derivation for Implication 3. For $i \neq j$, differentiating (8) in s_t gives the pairwise sensitivity in equation (9),

$$\beta_{ij}^{(h)} \equiv \frac{\partial \rho_{ij,t}^{(h)}}{\partial s_t} = a_i a_j.$$

Splitting each pair equally between its two firms and summing gives firm i's contribution to the aggregate in equation (10),

$$k_i^{(h)} = \frac{1}{2}\sum_{j \neq i} \beta_{ij}^{(h)} = \frac{1}{2} a_i \sum_{j \neq i} a_j,$$

and these contributions add up to the aggregate, $\sum_i k_i^{(h)} = \frac{1}{2}\sum_{i\neq j} a_i a_j$. The diagonal carries no information, since correlations satisfy $\rho_{ii,t}^{(h)} \equiv 1$ and hence $\partial\rho_{ii,t}^{(h)}/\partial s_t = 0$. The signs follow from the exposures: $\beta_{ij}^{(h)} = a_i a_j$ is positive for pairs whose exposures share a sign and negative for oppositely exposed pairs, and $k_i^{(h)}$ has the sign of $a_i \sum_{j\neq i} a_j$. Under broadly same-signed exposures $\sum_{j\neq i} a_j$ shares the common sign, so $k_i^{(h)}$ inherits the sign of a_i and most firms contribute in the same direction. If $a_i \geq 0$ for all firms, every pair sensitivity and every firm contribution is nonnegative.

B Fisher-Transformed Correlations

The empirical specifications use the Fisher transform $z(\rho) = \operatorname{arctanh}(\rho)$, which is strictly increasing on the open interval from -1 to 1. Because it is strictly increasing, it preserves the sign of every comparative static in the implications: these are claims about signs, and a monotone reparameterization of the dependent variable cannot change them. If the dependent variable is $z(\bar{\rho}_t^{(h)})$, then

$$\frac{\partial z(\bar{\rho}_t^{(h)})}{\partial s_t} = \frac{1}{1-\left(\bar{\rho}_t^{(h)}\right)^2}\frac{\partial \bar{\rho}_t^{(h)}}{\partial s_t},$$

which carries the same sign as the untransformed derivative in Implication 1. If the dependent variable instead averages transformed pairwise correlations, then

$$\frac{\partial z(\rho_{ij,t}^{(h)})}{\partial s_t} = \omega_{ij,t}^{(h)} a_i a_j, \qquad \omega_{ij,t}^{(h)} \equiv \frac{1}{1-\left(\rho_{ij,t}^{(h)}\right)^2} > 0.$$

Pairwise signs in Implication 3 are preserved because the weights are positive. For an aggregate average of transformed pairwise correlations, the derivative is a weighted sum of the exposure products. Positive weights preserve the sign of the unweighted sum automatically under the stronger same-signed condition $a_i \geq 0$ for all firms; under the weaker average-product condition, the same sign follows when the weights are close to constant. For monthly return correlations, which are far from ± 1, this weighting distinction is second order.

Online Appendix

C Additional Results

Table A1: Decomposition into within-direction concentration and cross-direction relatedness. This table decomposes total relatedness α_t into within-direction concentration R_t and cross-direction relatedness S_t. The dependent variable is the Fisher-transformed average pairwise return correlation at $t + 12$ in Panel (a) and the Fisher-transformed top-eigenvalue share of the implied correlation matrix at $t + 12$ in Panel (b). Columns (1)–(2) report total relatedness alone; columns (3)–(4) include standardized within-direction concentration and cross-direction relatedness separately. Baseline controls are uncertainty ($\ln(EPU)$, $\ln(VIX)$) and funding ($\ln(IPO + 1)$, $\ln(VC\ flow + 1)$); full controls add sentiment and TNIC breadth. The implied sign pattern ($\beta_R < 0$, $\beta_S > 0$) validates the construction behind the cross-direction share s_t. Standard errors are Newey–West with 12 monthly lags. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Panel (a): Average pairwise correlation at $t + 12$

	(1)	(2)	(3)	(4)
Total relatedness (α_t)	-0.043^{*}	-0.031		
	(0.023)	(0.021)		
Within-direction concentration (R_t)			-0.114^{***}	-0.162^{***}
			(0.036)	(0.049)
Cross-direction relatedness (S_t)			0.065^{**}	0.109^{***}
			(0.030)	(0.041)
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment + TNIC	No	Yes	No	Yes
Observations	277	277	277	277

Panel (b): Top-eigenvalue share at $t + 12$

	(1)	(2)	(3)	(4)
Total relatedness (α_t)	-0.044^{*}	-0.032		
	(0.023)	(0.021)		
Within-direction concentration (R_t)			-0.110^{***}	-0.155^{***}
			(0.035)	(0.048)
Cross-direction relatedness (S_t)			0.060^{**}	0.102^{**}
			(0.031)	(0.041)
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment + TNIC	No	Yes	No	Yes
Observations	277	277	277	277

Table A2: Technology-area funding controls. This table re-estimates the main specifications adding technology-area funding-flow controls for biotech, healthcare, financial services, energy, AI and data, software, consumer, hardware, and business services. Panel (a) uses monthly technology-area funding flows. Panel (b) uses trailing 12-month technology-area funding flows. The dependent variable is the Fisher-transformed average pairwise return correlation at $t+12$ in columns (1)–(2) and the Fisher-transformed top-eigenvalue share of the implied correlation matrix at $t+12$ in columns (3)–(4). The cross-direction share s_t and total relatedness α_t are standardized. Standard errors are Newey–West with 12 monthly lags. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Panel (a): Monthly technology-area funding flows

	Average pairwise correlation		Top-eigenvalue share	
	(1)	(2)	(3)	(4)
Cross-direction share (s_t)	0.046**	0.058**	0.044**	0.055**
	(0.019)	(0.024)	(0.019)	(0.024)
Total relatedness (α_t)	−0.021	−0.015	−0.022	−0.017
	(0.021)	(0.020)	(0.021)	(0.020)
Technology-area flows	Yes	Yes	Yes	Yes
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment + TNIC	No	Yes	No	Yes
Observations	277	277	277	277

Panel (b): Trailing 12-month technology-area funding flows

	Average pairwise correlation		Top-eigenvalue share	
	(1)	(2)	(3)	(4)
Cross-direction share (s_t)	0.162***	0.166***	0.158***	0.161***
	(0.038)	(0.038)	(0.037)	(0.038)
Total relatedness (α_t)	0.046	0.058	0.045	0.058
	(0.041)	(0.040)	(0.040)	(0.039)
Technology-area flows	Yes	Yes	Yes	Yes
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment + TNIC	No	Yes	No	Yes
Observations	277	277	277	277

Table A3: Alternative composition measure: the log relatedness ratio. This table repeats the main specifications of Tables 2 and 3 with the standardized log relatedness ratio $w_t = \ln(S_t/R_t) = \text{logit}(s_t)$ in place of the standardized cross-direction share s_t, alongside standardized total relatedness α_t. The dependent variable is the Fisher-transformed average pairwise return correlation at $t+12$ in Panel (a) and the Fisher-transformed top-eigenvalue share of the implied correlation matrix at $t+12$ in Panel (b). Column structure and controls follow the main tables. Standard errors are Newey–West with 12 monthly lags. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Panel (a): Average pairwise correlation at $t + 12$

	(1)	(2)	(3)	(4)
Log relatedness ratio (w_t)	0.047***	0.075***	0.068**	0.045***
	(0.016)	(0.021)	(0.028)	(0.016)
Total relatedness (α_t)	−0.021	−0.011	−0.016	0.004
	(0.022)	(0.021)	(0.023)	(0.021)
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment	No	Yes	Yes	No
Innovation control	No	TNIC	Patent	Tech Buzz
Observations	277	277	277	199

Panel (b): Top-eigenvalue share at $t + 12$

	(1)	(2)	(3)	(4)
Log relatedness ratio (w_t)	0.045***	0.071***	0.064**	0.042***
	(0.016)	(0.021)	(0.028)	(0.016)
Total relatedness (α_t)	−0.023	−0.013	−0.019	0.003
	(0.022)	(0.021)	(0.023)	(0.021)
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment	No	Yes	Yes	No
Innovation control	No	TNIC	Patent	Tech Buzz
Observations	277	277	277	199

Table A4: Longer category-similarity estimation windows. This table repeats the main specifications of Tables 2 and 3 using alternative windows to estimate the dynamic category-similarity matrix. The venture capital allocation shares are constructed from trailing 12-month funding flows, while the category-similarity matrix is estimated using 24-month and 36-month category co-assignment windows. The dependent variable is the Fisher-transformed average pairwise return correlation at $t+12$ in Panel (a) and the Fisher-transformed top-eigenvalue share of the implied correlation matrix at $t+12$ in Panel (b). Within each similarity-window block, the first column includes uncertainty and funding controls; columns (2)–(4) add additional controls. The cross-direction share s_t and total relatedness α_t are standardized. Uncertainty controls are $\ln(EPU)$ and $\ln(VIX)$; funding controls are $\ln(IPO+1)$ and $\ln(VC\ flow+1)$; the innovation control is TNIC breadth, patent value, or BNBC news attention (Tech Buzz), one at a time. Standard errors are Newey–West with 12 monthly lags and are reported in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Panel (a): Average pairwise correlation at $t+12$

	24-month similarity window				36-month similarity window			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Cross-direction share (s_t)	0.043*	0.052**	0.048*	0.026	0.072***	0.071***	0.071***	0.050***
	(0.025)	(0.021)	(0.026)	(0.018)	(0.021)	(0.022)	(0.025)	(0.014)
Total relatedness (α_t)	−0.017	−0.015	−0.019	0.003	0.007	0.008	0.005	0.019
	(0.024)	(0.026)	(0.027)	(0.024)	(0.021)	(0.022)	(0.023)	(0.020)
Uncertainty controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sentiment	No	Yes	Yes	No	No	Yes	Yes	No
Innovation control	No	TNIC	Patent	Tech Buzz	No	TNIC	Patent	Tech Buzz
Observations	265	265	265	187	253	253	253	175

Panel (b): Top-eigenvalue share at $t+12$

	24-month similarity window				36-month similarity window			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Cross-direction share (s_t)	0.042*	0.050**	0.046*	0.025	0.072***	0.071***	0.071***	0.050***
	(0.026)	(0.022)	(0.026)	(0.018)	(0.022)	(0.022)	(0.025)	(0.014)
Total relatedness (α_t)	−0.019	−0.016	−0.021	0.001	0.006	0.008	0.004	0.018
	(0.024)	(0.026)	(0.027)	(0.024)	(0.021)	(0.021)	(0.023)	(0.020)
Uncertainty controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sentiment	No	Yes	Yes	No	No	Yes	Yes	No
Innovation control	No	TNIC	Patent	Tech Buzz	No	TNIC	Patent	Tech Buzz
Observations	265	265	265	187	253	253	253	175

Table A5: Robustness to non-Fisher-transformed outcomes. This table repeats the main specifications of Tables 2 and 3 using the non-Fisher-transformed average pairwise return correlation at $t+12$ in Panel (a) and the non-Fisher-transformed top-eigenvalue share of the implied correlation matrix at $t+12$ in Panel (b). The cross-direction share s_t and total relatedness α_t are standardized. The first column includes uncertainty and funding controls; columns (2)–(4) add additional controls. Uncertainty controls are $\ln(EPU)$ and $\ln(VIX)$; funding controls are $\ln(IPO+1)$ and $\ln(VC\ flow+1)$; the innovation control is TNIC breadth, patent value, or BNBC news attention (Tech Buzz), one at a time. Standard errors are Newey–West with 12 monthly lags and are reported in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Panel (a): Average pairwise correlation at $t+12$

	(1)	(2)	(3)	(4)
Cross-direction share (s_t)	0.041***	0.065***	0.059**	0.039***
	(0.014)	(0.019)	(0.025)	(0.014)
Total relatedness (α_t)	−0.019	−0.010	−0.015	0.003
	(0.019)	(0.019)	(0.021)	(0.018)
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment	No	Yes	Yes	No
Innovation control	No	TNIC	Patent	Tech Buzz
Observations	277	277	277	199

Panel (b): Top-eigenvalue share at $t+12$

	(1)	(2)	(3)	(4)
Cross-direction share (s_t)	0.039***	0.061***	0.055**	0.036**
	(0.014)	(0.018)	(0.024)	(0.014)
Total relatedness (α_t)	−0.020	−0.012	−0.017	0.001
	(0.019)	(0.018)	(0.020)	(0.018)
Uncertainty controls	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes
Sentiment	No	Yes	Yes	No
Innovation control	No	TNIC	Patent	Tech Buzz
Observations	277	277	277	199

Table A6: Robustness to alternative Newey–West lag lengths. This table reports the first two specifications of Tables 2 and 3 using Newey–West standard errors with 12, 18, and 24 monthly lags. The 12-lag columns reproduce the baseline estimates reported in Tables 2 and 3 and are included here for comparison. The dependent variable is the Fisher-transformed average pairwise return correlation at $t+12$ in Panel (a) and the Fisher-transformed top-eigenvalue share of the implied correlation matrix at $t+12$ in Panel (b). The cross-direction share s_t and total relatedness α_t are standardized. Within each lag block, column (1) includes uncertainty and funding controls and column (2) adds sentiment and TNIC breadth. Standard errors are reported in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Panel (a): Average pairwise correlation at $t+12$

	12 lags		18 lags		24 lags	
	(1)	(2)	(1)	(2)	(1)	(2)
Cross-direction share (s_t)	0.047***	0.075***	0.047***	0.075***	0.047***	0.075***
	(0.016)	(0.021)	(0.017)	(0.022)	(0.018)	(0.023)
Total relatedness (α_t)	−0.020	−0.011	−0.020	−0.011	−0.020	−0.011
	(0.022)	(0.021)	(0.022)	(0.023)	(0.022)	(0.023)
Uncertainty controls	Yes	Yes	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes	Yes	Yes
Sentiment	No	Yes	No	Yes	No	Yes
Innovation control	No	TNIC	No	TNIC	No	TNIC
Observations	277	277	277	277	277	277

Panel (b): Top-eigenvalue share at $t+12$

	12 lags		18 lags		24 lags	
	(1)	(2)	(1)	(2)	(1)	(2)
Cross-direction share (s_t)	0.045***	0.071***	0.045**	0.071***	0.045**	0.071***
	(0.016)	(0.021)	(0.017)	(0.022)	(0.018)	(0.023)
Total relatedness (α_t)	−0.023	−0.013	−0.023	−0.013	−0.023	−0.013
	(0.022)	(0.021)	(0.023)	(0.023)	(0.023)	(0.023)
Uncertainty controls	Yes	Yes	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes	Yes	Yes
Sentiment	No	Yes	No	Yes	No	Yes
Innovation control	No	TNIC	No	TNIC	No	TNIC
Observations	277	277	277	277	277	277

Table A7: Major market and innovation episode leave-outs. This table repeats the first two specifications of Tables 2 and 3 in the full sample and after excluding, one window at a time, formation months in 2000–2003, 2008–2009, 2020–2021, and 2022–2023. The dependent variable is the Fisher-transformed average pairwise return correlation at $t+12$ in Panel (a) and the Fisher-transformed top-eigenvalue share of the implied correlation matrix at $t+12$ in Panel (b). The cross-direction share s_t and total relatedness α_t are standardized. Within each full-sample or excluded-period block, the first column includes uncertainty and funding controls and the second column adds sentiment and TNIC breadth. All specifications are estimated by OLS, with standard errors reported in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Panel (a): Average pairwise correlation at $t+12$

	Full sample		Excl. 2000–2003		Excl. 2008–2009		Excl. 2020–2021		Excl. 2022–2023	
Cross-direction share (s_t)	0.047***	0.075***	0.117***	0.137***	0.039***	0.064***	0.051***	0.075***	0.042***	0.076***
	(0.008)	(0.011)	(0.020)	(0.022)	(0.008)	(0.012)	(0.008)	(0.011)	(0.009)	(0.013)
Total relatedness (α_t)	−0.020	−0.011	0.003	0.005	−0.026*	−0.017	−0.012	−0.002	−0.018	−0.014
	(0.013)	(0.013)	(0.016)	(0.015)	(0.013)	(0.013)	(0.014)	(0.013)	(0.014)	(0.014)
Uncertainty controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sentiment	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Innovation control	No	TNIC	No	TNIC	No	TNIC	No	TNIC	No	TNIC
Observations	277	277	240	240	253	253	253	253	253	253

Panel (b): Top-eigenvalue share at $t+12$

	Full sample		Excl. 2000–2003		Excl. 2008–2009		Excl. 2020–2021		Excl. 2022–2023	
Cross-direction share (s_t)	0.045***	0.071***	0.120***	0.137***	0.037***	0.061***	0.049***	0.071***	0.039***	0.071***
	(0.008)	(0.011)	(0.020)	(0.022)	(0.008)	(0.011)	(0.008)	(0.011)	(0.009)	(0.013)
Total relatedness (α_t)	−0.023*	−0.013	0.003	0.004	−0.029**	−0.019	−0.014	−0.005	−0.020	−0.016
	(0.013)	(0.013)	(0.015)	(0.015)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)
Uncertainty controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Funding controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sentiment	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Innovation control	No	TNIC	No	TNIC	No	TNIC	No	TNIC	No	TNIC
Observations	277	277	240	240	253	253	253	253	253	253

Table A8: Matrix-label placebo regressions. This table reports matrix-label placebo tests for the main predictive specifications. For each placebo draw, we randomly permute the mapping between industry categories and the category-similarity matrix while holding fixed monthly venture capital funding shares across categories. We then reconstruct the placebo cross-direction share and re-estimate the specifications in Tables 2 and 3. The table reports the observed coefficient on the standardized cross-direction share and the distribution of corresponding placebo coefficients across 1,000 relabelings. The one-sided placebo p-value is computed as $(1 + \#\{\beta_{\text{placebo}} \geq \beta_{\text{observed}}\})/(1 + 1{,}000)$. All regressions use Newey–West standard errors with 12 monthly lags.

Specification	Observed coefficient	Placebo mean	Placebo SD	p5	Median	p95	p-value	Placebos	N
Panel (a): Average pairwise correlation at $t+12$									
Baseline controls	0.047	0.001	0.009	-0.014	0.001	0.016	0.001	1,000	277
Full controls with TNIC	0.075	0.000	0.009	-0.016	0.000	0.015	0.001	1,000	277
Full controls with patent	0.068	0.000	0.009	-0.016	-0.001	0.015	0.001	1,000	277
Full controls with Tech Buzz	0.046	-0.008	0.011	-0.026	-0.008	0.011	0.001	1,000	199
Panel (b): Top-eigenvalue share at $t+12$									
Baseline controls	0.045	0.001	0.009	-0.014	0.001	0.016	0.001	1,000	277
Full controls with TNIC	0.071	0.000	0.009	-0.016	0.000	0.015	0.001	1,000	277
Full controls with patent	0.065	0.000	0.009	-0.016	0.000	0.015	0.001	1,000	277
Full controls with Tech Buzz	0.043	-0.007	0.011	-0.025	-0.007	0.011	0.001	1,000	199

Figure A1: Observed cross-direction share versus matrix-label placebo shares. This figure compares the observed cross-direction share with the distribution of cross-direction shares obtained from 1,000 matrix-label placebo relabelings. In each placebo draw, the mapping between industry categories and the category-similarity matrix is randomly permuted while monthly venture capital funding shares across categories are held fixed. The shaded region shows the 5th–95th percentile range of placebo shares, the solid line shows the placebo median, and the dashed red line shows the observed cross-direction share.

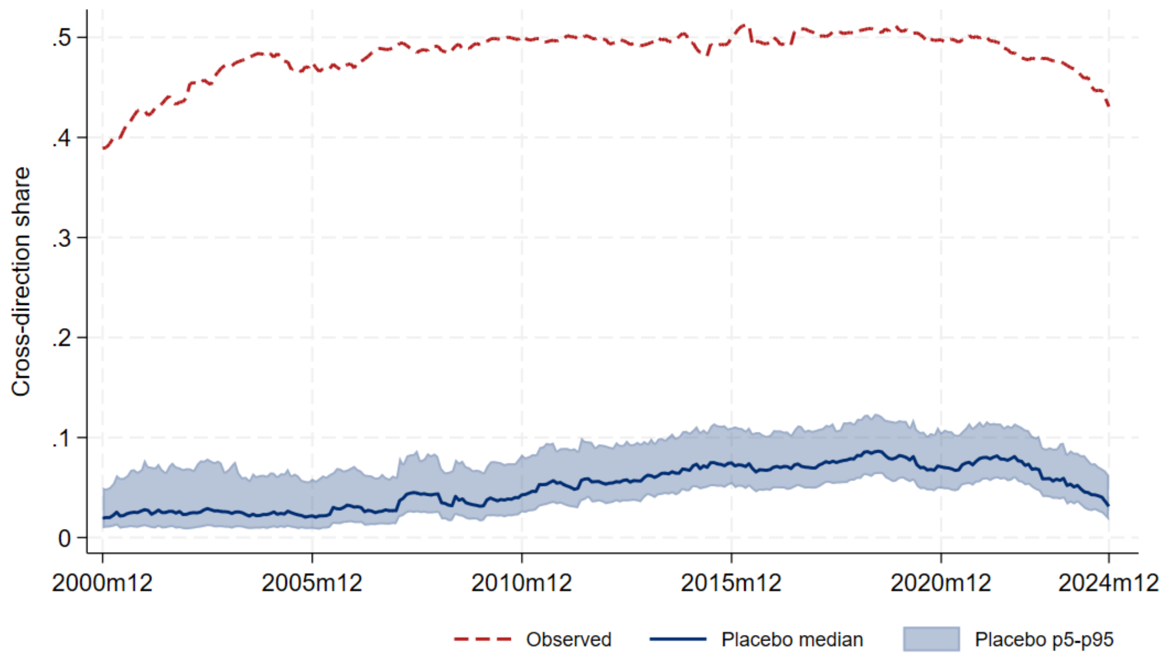
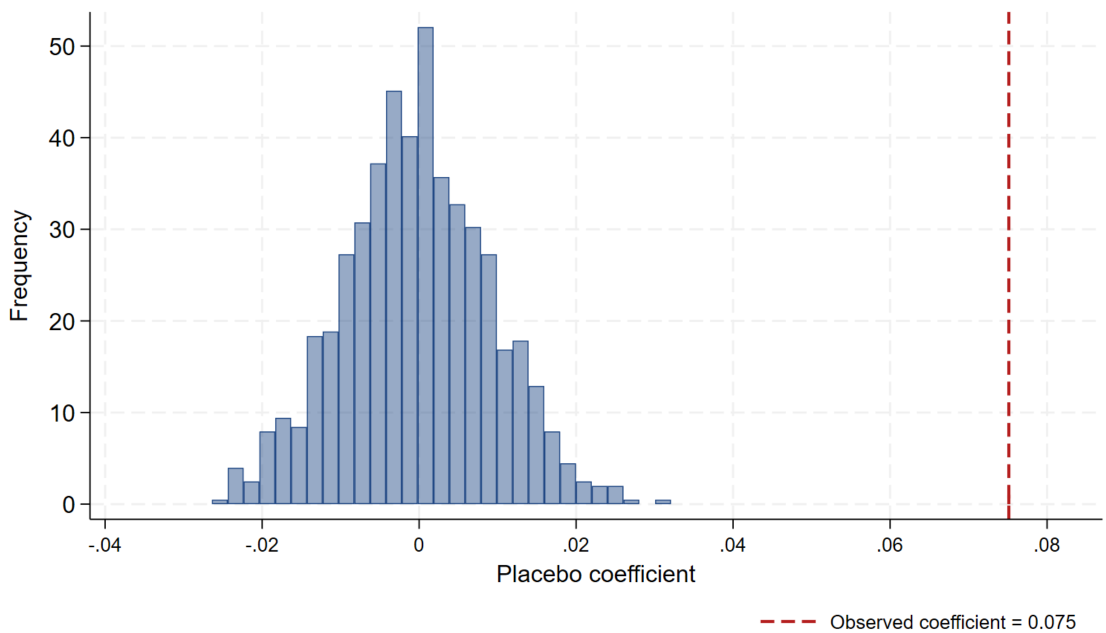


Figure A2: Placebo coefficient distributions: full-control specifications. This figure shows the distribution of placebo coefficients from 1,000 matrix-label relabelings for the full-control specifications. For each placebo draw, the mapping between industry categories and the category-similarity matrix is randomly permuted while monthly venture capital funding shares across categories are held fixed. The dashed red line marks the observed coefficient from the corresponding main regression.

Panel (a): Average pairwise correlation at $t + 12$



Panel (b): Top-eigenvalue share at $t + 12$