Venture Capital and Customer Capital*

Martin Aragoneses
INSEAD

Young Soo Jang
Penn State

Amir Sufi
Chicago Booth & NBER

September 2026

Preliminary for feedback & conference submissions
Please Do Not Circulate
Public draft available October 2026

Abstract

Using data sets on the intangible investments, financing, cash flows, and returns of small, young, private firms in the United States, this study explores the unique function of venture capital (VC) in the economy relative to banks, finance companies, private equity, and private credit. Relative to other financial intermediaries, venture capital is special in its focus on financing customer capital intensive firms for which trademarks, sales and marketing investments, and a large sales force are crucial for the business. The VC focus on customer capital intensive firms is also evident among the biggest winners that ultimately go public. Firms tilt investment toward customer capital leading up to the initial VC investment, after which their scale accelerates rapidly. Analysis of detailed investment cash flow data provides an explanation for why venture capital specializes in customer capital: Financing customer capital requires tolerance for downside risk and patience for delayed value realization in pursuit of extreme upside potential, as characterized by power law returns. This return profile fits customer capital investments more closely than investments in knowledge capital.

*We thank Xiangyu Chen, Tanvi Jindal, and Oleg Telegin who provided excellent research assistance, Jessica Jeffers (discussant), and participants at the INSEAD Finance Symposium for helpful comments. We are also grateful to the Institute for Private Capital and MSCI for their support and for granting us access to the MSCI Private Capital Universe database, and to Wendy Hu for her invaluable assistance with the data. Contact information: Aragoneses: martin.aragoneses@insead.edu; Jang: ykj5180@psu.edu; Sufi: amir.sufi@chicagobooth.edu.

You can't wait for users to come to you. You have to go out and get them.

— Paul Graham, Y Combinator

Venture capitalists finance some of the most influential businesses in the economy (e.g., Kaplan and Lerner (2010), Puri and Zarutskie (2012)). How are venture capitalists special relative to other financial intermediaries such as banks or private equity firms? The prevailing view in the existing literature is that venture capital plays a unique and transformational role in financing early-stage firms engaged in research and development and technological innovation (e.g., Gompers and Lerner (1999); Hall and Lerner (2010); Kerr and Nanda (2015); Lerner and Nanda (2020); Howell, Lerner, Nanda and Townsend (2020)).

This study presents a complementary view: venture capital is special relative to other financial intermediaries because of its focus on customer capital intensive businesses. Firms spend substantial resources building a customer base; they hire a sales force, collect data on potential and existing customers, file trademarks to protect their product, and advertise. Theoretical and empirical research shows that customer capital investment is on the same order of magnitude as capital expenditures, and such investment builds a customer base that is valuable (e.g., Larkin (2013), Gourio and Rudanko (2014); Belo, Lin and Vitorino (2014); He, Mostrom and Sufi (2026)). Furthermore, over the last decade, industries that invest most heavily in customer capital, such as information and business services, have increased their share of overall enterprise value.

The central point of this study is that venture capital–more than any other type of financial intermediary–specializes in the financing of customer capital intensive firms. They do so because investments in customer capital intensive firms, even more than investments in knowledge capital intensive firms, require an investor to tolerate substantial downside risk and prolonged periods of low cash flow in pursuit of the outsized long-run returns generated by a small number of big winners. This is precisely the return profile associated with venture capital.

Analyzing venture capital requires data sets that measure the intangible investments, financing, cash flows, and returns of small, young, private firms. We use three data sets to do so. The first covers the universe of all U.S. businesses, their financing, and measures of their customer and knowledge capital intensity as proxied by workforce composition and filings for patents and trademarks. The second focuses on firms before and through an initial public offering (IPO), and it contains information on whether the firm received VC financing and the firm's sales and marketing and research and development expenses. The third measures cash flow and return characteristics for investments in firms made by a broad range of private capital intermediaries, including venture

capital, private equity, and private credit.

Customer capital intensity exhibits substantial variation across industries (He, Mostrom and Sufi (2026)), and the analysis here starts by showing that the presence of venture capitalists in an industry is strongly positively correlated with this intensity. Specifically, industries in which venture capital has a larger presence have higher sales and marketing expense to revenue ratios, a higher share of customer capital to total capital, a higher share of their wage bill dedicated to sales and marketing employees, and a higher share of companies that have obtained a trademark.

The fact that customer capital intensive industries have a large share of firms financed with venture capital is in part due to the fact that knowledge capital intensity and customer capital intensity are positively correlated across industries. Indeed, the VC share is highest in industries such as software publishing and high-tech manufacturing which invest heavily in both knowledge and customer capital. However, VC presence is surprisingly high even in high customer capital intensity but low knowledge capital intensity industries, such as beverage and clothing manufacturing. Furthermore, VC investment is limited in knowledge capital intensive industries that do not invest heavily in customer capital. The industry level evidence hints that it is perhaps the presence of customer capital that is the necessary ingredient in attracting VC investment.

Moving to the firm level, we show that, even within narrowly-defined industry-geography-age bins, firms financed with venture capital have a significantly higher share of their wage bill dedicated to sales and marketing workers and a higher propensity to have filed for a trademark. In terms of magnitudes, VC-financed firms have a wage bill in sales and marketing that is 4 percentage points higher (on a mean of 25%), and a 39pp higher propensity to have filed a trademark (on a mean of 12%). The marginal effects associated with venture capital are significantly larger than any other type of financial intermediary, including private equity, private credit, banks, and finance companies. In fact, venture capital is the only intermediary associated with a higher share of the wage bill dedicated to sales and marketing.

Ultimately, the macroeconomic impact of this customer capital focus of VC-financed firms depends on the big winners that go on to become some of the largest firms in the economy. The analysis shows that firms financed by venture capital that go public have significantly higher sales and marketing expenses to revenue ratios just prior to the IPO. Even with the inclusion of industry fixed effects, VC-financed firms have a sales and marketing to revenue ratio that is 15 percentage points higher than non-VC-financed firms. VC-backed firms that IPO have significantly higher enterprise value to physical capital ratios in the years after the IPO, which is consistent with a more

intangible capital intensive business model (e.g., Crouzet and Eberly (2023)).

How is VC financing associated with the dynamics of customer capital investment and business scale for firms receiving the investment? We construct two research designs to evaluate the dynamics of these firms around the initial VC investment: one is based on a matched control sample of firms never receiving a VC investment and the other is based on a control sample of firms that receive a VC investment later in the sample.

A robust finding holds in both designs: firms invest heavily in customer capital leading up to the initial VC investment. There is a rise in the share of workers working in sales and marketing in the two years prior to the investment, and there is a substantial increase in the probability of filing for a trademark in the months just prior to receiving the initial VC investment. The relative share of workers working in sales and marketing for firms receiving a VC investment does not rise after the investment, which suggests that the focus on customer capital by the firms occurs even before they receive VC financing.

While the composition of the workforce tilts toward sales and marketing prior to the VC investment, the scale of the business accelerates rapidly after the investment. Both research designs point to a dramatic relative growth of VC-financed businesses in the three years after the investment of 200 to 300% relative to the control group. The overall results on dynamics suggest that firms receiving a VC investment invest heavily in customer capital just prior to the investment, and then they grow rapidly afterward. The results are consistent with the view that firms must establish customer traction and "product market fit" to attract an initial VC investment, and then VC financing helps scale the business afterward.

Why might venture capital specialize in financing customer capital? To answer this question, the analysis focuses on private capital investment returns in the MSCI Private Capital Universe (formerly known as Burgiss). This investment-level data set allows us to measure the return characteristics of investments across investor type (e.g., venture capital, private equity, and private credit) and the return characteristics of investments across industries with different knowledge and customer capital intensities. Looking at the distribution of returns by investor type, the analysis shows that, despite similar mean and variance, VC investments exhibit greater downside risk in the lower tail and greater upside potential in the upper tail, consistent with the heavy-tailed power law distribution typically associated with VC investments (Cochrane, 2005; Puri and Zarutskie, 2012; Kerr, Nanda and Rhodes-Kropf, 2014b; Korteweg and Nagel, 2016; Brown, Harris, Hu, Jenkinson, Kaplan and Robinson, 2020). Moreover, VC investments have lower current cash distributions relative to un-

realized future value, demonstrating venture capitalists' greater tolerance for illiquidity or slower value realization (Sahlman, 1990; Gompers, 1995; Hellmann, Montag and Tå g, 2024).

The MSCI data set shows that investments in customer capital intensive industries exhibit these exact same characteristics. Specifically, they are more likely to have a zero distribution per investment dollar ratio (DPI), a lower ratio of DPI to total value per investment ratio (TVPI), and a higher probability of having a TVPI in the top 1% of the distribution. Investment level regressions show that these results are robust across the firm age distribution. Perhaps surprisingly, investments in customer capital industries resemble the typical VC return profile more strongly than investment in knowledge capital intensive industries. Overall, the results show that investments in customer capital intensive industries, even more than investments in knowledge capital industries, exhibit return characteristics that match what is typically associated with VC investments.

Related Literature This paper fits within the broader literature that seeks to explain the unique role played by venture capital, in particular, in the financing of intangible capital in an economy. As mentioned above, a primary emphasis of this literature is on the role venture capital plays in financing young, innovative, and technologically intensive firms, with particular attention to R&D, patents, and high-tech industries (Gompers and Lerner (1999), Kortum and Lerner (2000), Hall and Lerner (2010), Kerr and Nanda (2015), Lerner and Nanda (2020), and Howell, Lerner, Nanda and Townsend (2020)).[1] Many of the results of this paper support this view: VC-financed firms have higher wage bill shares in technical occupations, have a significantly higher propensity to apply for patents, and are in R&D intensive industries. However, the results here emphasize a complementary role of venture capital in financing businesses that rely heavily on customer capital to generate value.

Existing research in venture capital emphasizes the importance of customer acquisition and

[1] A longstanding literature highlights the important role venture capitalists play in adding value to innovative firms by providing "smart" money (Sorensen, 2007), often drawing on their own entrepreneurial experience (e.g. Ewens and Rhodes-Kropf (2015); Gompers and Mukharlyamov (2022); Jackson and Strebulaev (2026)), devote substantial resources to screening and due diligence (e.g., Bernstein, Korteweg and Laws (2017); Gompers, Gornall, Kaplan and Strebulaev (2020); Gonzalez-Uribe, Reyes and Yin (2026); Jang and Kaplan (2026); Lyonnet and Stern (2026)), and mitigate information asymmetries through sophisticated contracting (e.g., Hellmann (1998); Kaplan and Strömberg (2003); Kaplan and Strömberg (2004); Ewens, Gorbenko and Korteweg (2022)) and active monitoring (e.g., Gompers (1995); Lerner (1995); Bernstein, Giroud and Townsend (2016); Gompers, Gornall, Kaplan and Strebulaev (2020)). Venture capitalists also frequently syndicate and stage financing: syndication facilitates risk sharing and pools complementary expertise (e.g., Hochberg, Ljungqvist and Lu (2007)), while staged financing preserves entrepreneurial incentives and accommodates the prolonged J-curves of innovative firms (e.g., Sahlman (1990); Gompers (1995); Gornall and Strebulaev (2020); Hellmann, Montag and Tå g (2024)). VC-backed firms also benefit from the sharing of innovation resources across companies within VC portfolios (González-Uribe, 2020).

product market fit. For example, Hellmann and Puri (2000) and Hellmann and Puri (2002) provided early evidence that VC investments are associated with reduced time to bring a product to market and a rise in the presence of a marketing vice president (VP). Another key study is De Vries, Pennings, Block and Fisch (2017), who show that VC-financed businesses are more likely than other startups to file a trademark application relative to a patent application. This study shares this focus on the importance of customer capital for venture capital, but examines it across the broader universe of firms and relative to a broader set of intermediaries. While VC financing may be associated with a faster time it takes to bring a product to market, the results here show that investing in customer capital significantly delays cash flow distributions to investors. This makes customer capital investment more risky and therefore more amenable for relatively risk-tolerant and patient investors like venture capitalists.

This study also contributes to the literature that emphasizes the importance of customer capital in venture capitalists' investment selection. Foundational evidence on this comes from Kaplan and Strömberg (2004): 69% venture capitalists surveyed in their article report that market size and growth are a reason to invest, and 30% emphasize a high likelihood of customer adoption as a reason to invest. Another key study is Gompers, Gornall, Kaplan and Strebulaev (2020), who use survey evidence to show a combination of both selection and treatment when it comes to venture capitalists and customer acquisition. VCs often emphasize the product and the market as important selection criteria for investing, and they also emphasize that they help portfolio companies by connecting then to potential customers. For trademarks, Block, De Vries, Schumann and Sandner (2014) examine how the number of trademarks affects the valuation of a company by a venture capitalist, and they find an inverted U-shape relationship.

The finding here that firms significantly tilt investment toward customer capital and file trademarks immediately before obtaining VC financing provides complementary evidence that customer capital is an important selection criterion in VC investment decisions. Furthermore, we find that the primary post-investment VC effect is to scale the business, while the share of sales and marketing employment remains stable. This finding is consistent with Kaplan, Sensoy and Strömberg (2009), who trace firm characteristics from the early business-plan stage through IPO and into the public-company stage for VC-financed firms and show that most firms continue to serve similar customers throughout these stages. Taken together, our findings suggest that venture capitalists select firms based on customer-related characteristics and subsequently help them scale.

This work also relates to several studies at the intersection of macroeconomics and finance

that speak to firm growth and investments firms make and how they are financed, especially early in the firm life cycle and when they involve investments in intangible capital, (e.g., McGrattan and Prescott (2010); Midrigan and Xu (2014); Gutiérrez and Philippon (2016); Peters and Taylor (2017); Falato, Kadyrzhanova, Sim and Steri (2022); Hoberg and Maksimovic (2022); Crouzet and Eberly (2023); Belo, Li, Salomao and Vitorino (2024); Ottonello and Winberry (2024); Crouzet, Eberly, Eisfeldt and Papanikolaou (2025); Aragoneses, Hansman, Licher and Nanda (2026); Cai, Prat and Yu (2026)). This study shows that financial intermediaries differ in their propensity to finance different types of intangible capital, with venture capital targeting customer capital intensive industries in particular.

1 Customer Capital: Background

Customer capital reflects the value of a customer base to a business. Given this value, firms invest in customer capital through activities such as building a brand, trademarking their products, hiring a sales force, gathering data on customers, and advertising. Early studies exploring the value of customer capital include Belo, Lin and Vitorino (2014); Bronnenberg, Dubé and Gentzkow (2012); Gourio and Rudanko (2014); and Larkin (2013). A wave of recent studies further explores this concept in both empirical and theoretical settings (e.g., Arellano-Bover, Bussotti, Paradisi and Wu (2026); Azinovic-Yang, Kepler, Speros and Stewart (2026); Chiavari (2026); He, Mostrom and Sufi (2026); He (2025); Hubmer and Nord (2025); Ignaszak and Sedláček (2026); Pearce and Wu (2025)).

There is no canonical model nor unified explanation in the literature for the source of this value. Search frictions have been modeled carefully (e.g., Gourio and Rudanko (2014)), but network effects in demand (e.g., He, Mostrom and Sufi (2026)), consumer behavioral biases (e.g., Della Vigna and Malmendier (2006); Einav, Klopack and Mahoney (2025)), and loyalty from brands (e.g., Bronnenberg, Dubé and Gentzkow (2012)) all find support in the data. The basic economic argument is quite robust to these various forces: firms earn rents from their customers once they are acquired, and they therefore are willing to spend substantial resources on customer acquisition.

Both the dynamic model of Gourio and Rudanko (2014) and the static model of Bond, Hashemi, Kaplan and Zoch (2021) point to the sales and marketing expense to revenue ratio as the appropriate measure of the extent to which a firm spends resources on customer acquisition. He, Mostrom and Sufi (2026) measure this variable for publicly-traded U.S. companies from 1997 to 2022, and they

show that firms spend substantial resources on customer acquisition.[2] Furthermore, He, Mostrom and Sufi (2026) show that these investments translate into a customer base that boosts value.

In the sample of private firms examined here, neither sales and marketing expense nor revenue are well measured. As a result, this study primarily relies on three measures of customer capital investment at the firm level: the ratio of total firm-level employment in sales and marketing (L^{SM}/L) roles, the ratio of the total wage bill in sales and marketing (WL^{SM}/WL) and the filing of trademarks. In the sample of IPOs described in Section 4, the sales and marketing to revenue ratio and the wage bill share in sales and marketing are strongly positively correlated. See Figure A1 in the appendix. This gives us confidence that the wage bill share is a good proxy for the sales and marketing to revenue ratio. At the industry level, the analysis also uses the median share of total capital in sales and marketing capital (K^{SM}/K) and the median ratio of sales and marketing expense to total revenue (I^{SM}/R). These measures are estimated on the set of publicly traded firms from 1997 to 2022, as in He, Mostrom and Sufi (2026).

We also use USPTO data to measure intangible investment among private and public firms. We first use patent filings as a measure of investment in knowledge capital. Supporting this interpretation, Kogan, Papanikolaou, Seru and Stoffman (2017) show that the economic value of patented innovations is associated with firm growth, resource reallocation, and creative destruction. We also then use trademark filings as a measure of investment in customer capital and product-market development. As stated in the introduction of Dinlersoz, Goldschlag, Myers and Zolas (2018), one of the first studies to combine the USPTO data with Census micro data: "Attracting consumers and retaining them as loyal customers is critical for a firm's survival and growth. Among the many ways of building and protecting a loyal customer base, trademarks are unique." Furthermore, Desai, Gavrilova, Silva and Soares (2022) show that trademarks are associated with subsequent product launches, investment, employment, profitability, and market share growth. The existing literature on trademarks supports the view that a trademark filing represents an investment in a customer base.[3]

[2] As an example, of the top 5 firms by market capitalization in the United States as of September 2025, 3 break out sales and marketing expense explicitly: Alphabet, Microsoft, and Amazon. Together, these three companies spent $102 billion on sales and marketing in 2025, which is on average 7.7% of revenue across the three firms.

[3] The existing research also uses trademarks as a proxy for branding innovation (e.g., Flikkema, Castaldi, De Man and Seip (2019)) and entrepreneurial quality (e.g., Guzman and Stern (2020)). This points to a complementarity between customer and knowledge capital that we explore below.

2 Data on Intangible Investment and Financing

There are three main data sets used in the empirical analysis: (1) a data set covering the universe of U.S. business in the United States from 2010 to 2022 based on NETS, PitchBook, Capital IQ, UCC filings, USPTO, and Revelio data, (2) a data set covering firms that undertake an initial public offering (IPO) at some point between 2010 and 2023, and (3) a data set based on investment-level information from MSCI. This section discusses the first of these data sets in detail, whereas discussion of the other two data sets is below where analysis of these data sets occurs.

2.1 NETS, PitchBook, Capital IQ, and UCC filings

A core contribution of this study is the construction of a data set that combines measures of the customer and knowledge capital intensity of U.S. firms with information on the sources of financing that these firms receive.

The data set covering the universe of U.S. businesses begins with the data set constructed in Jang, Kim, Sufi and Chen (2025). The backbone of the data set is the National Establishment Time Series (NETS) database, compiled by Walls & Associates from Dun & Bradstreet records. NETS provides comprehensive coverage of business establishments across the United States. The data set includes internal firm organizational structure that allows for the construction of a firm-level data set from the underlying establishment-level data. All analysis conducted in this study is at the firm-level. We follow similar sample restrictions as in Jang, Kim, Sufi and Chen (2025): we exclude firms owned by a foreign entity or operating in government (3-digit NAICS 921–928), religious and non-profit organizations (3-digit NAICS 813), educational institutions (3-digit NAICS 611), financial services (2-digit NAICS 52), or real estate and leasing (2-digit NAICS 53). This data set covers information on firms from 1990 to 2022, and much of the discussion below focuses on the cross-section of the data as of 2022, the last year of the sample.

While Jang, Kim, Sufi and Chen (2025) focus on middle market firms, defined as those between 21 and 1000 employees, the focus in this study is on smaller firms for which VC financing is crucial. As such, for the cross-sectional analysis below, the data set is isolated to all private firms with at least 1 employee and no more than 100 employees. As in Jang, Kim, Sufi and Chen (2025), the core data set here is augmented with information from PitchBook, Capital IQ, and UCC filings, which allows us to measure various sources of financing the firm receives (e.g., venture capital, private

equity, private credit, banks, and finance companies). Please see Jang, Kim, Sufi and Chen (2025) for more details on the construction of this data set.[4]

There are well known differences between the NETS data set and the official business data sets from the Census. Barnatchez, Crane and Decker (2017) and Crane and Decker (2020) discuss these issues in detail. Data on firm revenue from NETS are often unreliable due to high rates of imputation and they are therefore excluded from our analysis. Employment data, however, are generally reliable when analyzed using employment bins or ranges. As noted by Crane and Decker (2020), NETS is "reasonably representative of U.S. business activity in the static cross section," with strong correlations to official data in terms of size, industry, and geography, particularly when smaller establishments are excluded.

One issue raised by the analysis here is for small firms. Many venture-capital financed firms are extremely small, sometimes with only the founder as an employee. As discussed in Crane and Decker (2020), the NETS data set does not distinguish between non-employer businesses, which are mostly sole proprietorships with no payroll employees, and employer businesses. In contrast, official Census statistics on the number of U.S. businesses makes this distinction. As such, NETS has far more tiny "firms" relative to the Census. The gap is sizable in our sample, where we have 25 million firms with fewer than 5 employees as of 2022 whereas the Census has only 4 million. We should note from the outset that we believe that even sole proprieterships receiving a VC investment should be included in the analysis, which is an advantage of having the full NETS sample.

To explore this issue, Table A1 in the appendix shows information on the number and fraction of firms as of 2022 in the core data set that receive a VC investment according to either PitchBook or Capital IQ within the last five years. For this table, the size thresholds used by the Census are reported to make comparisons easy. In total, almost 50 thousand firms in the NETS universe with between 1 and 100 employees as of 2022 receive a VC investment at some point between 2018 and 2022. This is a lower bound given that it relies on PitchBook and Capital IQ coverage, and a successful match to NETS.

As a fraction of the firms in NETS, this represents only 0.17%. However, if the denominator is measured using the Census Statistics of U.S. Businesses, the fraction rises to 0.78%. This is the more accurate statistic given that all VC-financed firms at least aspire to be employer businesses rather than sole proprieterships. Table A1 also shows the fraction of VC-financed firms by size. As

[4]Given the focus of this study on small firms, we also implement a technique to remove establishments in NETS that are unlikely to be actual businesses. This technique is described in more detail in Section D in the appendix.

it shows, 32 thousand of the 50 thousand firms have less than five employees. Most of the firms financed by venture capital are extremely small.

The most comparable statistic in the literature comes from Puri and Zarutskie (2012), who match VentureSource and VentureXpert data on VC investments to the Census Longitudinal Business Data set. Over the 2001 to 2005 period, they find that 0.16% of firms have ever received a VC investment. This is substantially below our estimate of 0.78%, which could reflect the larger presence of venture capital over time or improved coverage of VC investments from sources such as PitchBook and Capital IQ.[5] As of 2022, we find that about 50 thousand firms received a VC investment at some point between 2018 to 2022. One comparable statistic comes from the PitchBook-NVCA Venture Monitor which reports that almost 18 thousand unique US companies received a VC investment in 2022 alone.

Table A2 in the appendix provides more details on the VC-financed firms in the final 2022 sample across the size, age, and geographical distribution. Almost 50% of VC-financed firms have five or fewer employees, and almost 50% are four years or younger in terms of age. Almost 18% of VC-financed firms are in the extremely large population counties of Manhattan county, San Francisco county, and Los Angeles county. However, the next three highest counties are Santa Clara (Silicon Valley), California; Middlesex (Cambridge), Massachusetts; and Travis (Austin), Texas, all known to be hotbeds of VC activity.

2.2 Patents, Trademarks, and Revelio

A central goal of the analysis is to examine how VC financing is related to measures of customer and knowledge capital investment at the firm level. To this end, the core data set is augmented with information on patent and trademark filings from the United States Patent and Trademark Office (USPTO), and information on the composition of the workforce of companies from Revelio. We follow Aghion, Aragoneses, Bergeaud and Ughi (2026), who develop a linking procedure to match USPTO data to firm-, establishment-, and worker-level data to study the interplay between VC activity and broadly defined measures of innovation.

The patents data set comes from PatentsView, which restructures USPTO patent records and provides disambiguated company identifiers together with patent, location, date, and technology-classification information (United States Patent and Trademark Office, 2026). For the cross-sectional

[5] Garfinkel, Mayer, Strebulaev and Yimfor (2025) compare Pitchbook with other VC databases, including Crunchbase and VentureXpert, and find that Pitchbook offers the most comprehensive coverage of entrepreneurial financing.

analyses, we construct cumulative patent counts and an indicator for whether a firm has any patents. For the dynamic analyses, we use annual and monthly filing-based measures. The USPTO Trademark Case Files Dataset contains application and registration case histories, including owner records, trademark characteristics, classifications, prosecution events, and renewal histories (Graham et al., 2013; United States Patent and Trademark Office, 2024). We construct cumulative trademark counts and an any-trademark indicator, together with annual and monthly filing-based measures for the dynamic analyses. Appendix Sections A and B describe in detail the match of patents and trademarks data to NETS using DUNS establishments identifiers in separate crosswalks.

The workforce composition measures come from Revelio Labs, which builds employment records from the near-universe of publicly available online professional profiles and resumes, primarily those on LinkedIn. The unit of observation in the Revelio data is a position: a worker's employment spell at a company, with a start month, an end month, a job title, a sampling weight, and a predicted salary. Revelio also maps each raw job title into a standardized role taxonomy, which is available at several levels of granularity, ranging from thousands of detailed roles to roughly ten broad occupational groups. We use the coarsest level, which Revelio labels `k10`.[6]

We classify a position as sales and marketing if it is assigned to the "Sales and Marketing" group, and as technical if it is assigned to the "Engineer," "Software Engineer," "Technician," or "Project and IT Specialist" groups. For the cross-sectional analysis, we take a snapshot of all positions active in December 2022 and aggregate them to the firm level using the sampling weights.

Revelio companies are linked to NETS firms using a multi-stage matching procedure based on company names, industries, and addresses. Appendix Section C describes the Revelio data and the construction and properties of the match in more detail. Overall, 4.4% of the firms in the NETS sample are matched to the Revelio data. This match probability is low in part because the NETS sample includes millions of very small businesses, many of which are sole proprietorships. Larger firms are much more likely to match, so the match probability is 12.4% when firms are weighted by the total number of employees. Almost 40% of VC-financed firms in the NETS sample are matched to Revelio; the matched sample captures a large fraction of VC-financed firms.

Table A3 in the appendix presents regressions relating match probability to various factors. The strongest determinant of a match is firm size. A firm with between 51 and 100 employees is 36pp more likely to be matched with Revelio compared to a firm with only 1 employee. Older firms are also significantly more likely to be matched.

[6] Appendix Table A7 lists the complete set of `k10` groups and how each is classified in this paper.

Firms that have received financing from either a venture capitalist or a private equity firm are 30pp and 27pp more likely to be matched to Revelio, which is higher than other types of financing. This could be due to the fact that employees at VC- and PE-financed firms experience greater benefits from being on LinkedIn. It could also be viewed as a type of customer capital investment itself: having the firm and its employees on LinkedIn boosts the publicity of these companies.

A comparison of the VC-financed firms that are and are not matched to Revelio yields broadly similar patterns, as shown in Table A2 in the appendix. Larger firms and older VC-financed firms are more likely to be matched, as in the full sample. However, the coverage is still excellent for extremely young and small firms: 27% of VC-financed firms that have only one employee are in Revelio and 30% of VC-financed firms that are only one year old are in Revelio.

One main advantage of the patent and trademark data is that it does not depend on the match with Revelio. As shown in the appendix, all results using the patent and trademark data are qualitatively similar in the full and matched Revelio sample.[7] This gives us confidence that the qualitative patterns shown here would be robust to the full sample.

2.3 Summary statistics

Table 1 provides summary statistics on the sample of firms with between 1 and 100 employees in the NETS-Revelio matched data set as of 2022. The average firm has 8 employees and is 17 years old. The means when weighting by the number of employees are 31 and 29, respectively.

In the matched sample, 1.5% of the firms have received a VC investment at some point between 2018 and 2022. This is slightly higher than PE or private credit, but much lower than either banks or finance companies. Almost 18% of firms have received a bank loan, as measured by UCC filings, between 2018 and 2022. Banks and finance companies are still the dominant financiers of small firms in the U.S. economy, consistent with the importance of external debt for entrepreneurs (Robb and Robinson, 2014).

Almost 2% of firms with between 1 and 100 employees have filed for a patent at some point in their lifetime, and 12% file for a trademark. The weighted means are significantly higher: 4% and 21%, respectively. On average, about 25% of the wage bill is paid to sales and marketing employees and about 21% of the wage bill is paid to technical workers.

[7] See in particular Tables A4 and A5, and Figures A2, A3, and A4 in the appendix.

3 Cross-sectional Evidence

3.1 Industry-level evidence

Customer capital investment intensity varies significantly across industries, and this variation is persistent over time (He, Mostrom and Sufi (2026)). How is this variation related to industry-level variation in the presence of venture capital? To answer this question, the analysis uses industry-level averages from the the 2022 cross-sectional NETS-based data set. Specifically, Table 2 reports regressions from the following specification:

$$Y_j = \alpha + \beta \, \text{VCshare}_j + \varepsilon_j, \tag{1}$$

where Y_j are measures of customer and knowledge capital intensity in industry j, and VCshare_j is the share of firms receiving a VC investment within the past five years. The sample is once again the set of private firms that have between 1 and 100 employees.

Panel A of Table 2 shows a statistically significant and economically sizable positive relationship between customer capital intensity and the fraction of firms receiving an investment from a venture capitalist in the last five years. The first two measures come from He, Mostrom and Sufi (2026). These variables are calculated on the set of publicly traded firms from 1997 to 2022. The latter two measures come from the NETS-based data set discussed in Section 2.

The positive relationship is statistically robust across all four measures customer capital investment intensity. The left hand side variable in column 5 represents the first principal component of the four customer capital measures, which is standardized to have mean zero and standard deviation one. The regression coefficient in column 5 implies that a 1pp increase in the share of venture capital leads to a one-third standard deviation increase in the customer capital intensity of the industry. Panel A of Figure 1 illustrates this relationship through a bin-scatter: moving from the bottom quintile to the top quintile of VC distribution is almost a 5pp increase, which is associated with a rise in the customer capital intensity of the industry of almost 2 standard deviations.

Consistent with earlier research from Gompers and Lerner (1999), Panel B shows a strong positive relationship between measures of knowledge capital intensity and VC presence. The fraction of firms financed by a venture capitalist is positively related to three of the four measures of knowledge capital intensity. Column 5 uses the first principal component of the four measures of knowledge

capital intensity as a left hand side variable: the magnitude of the coefficient is even larger than for customer capital intensity. Panel B of Figure 1 shows the bin scatter associated with knowledge capital.

In column 6 of Table 2, the full NETS sample is used in place of the sample that has been matched with NETS. All variables are available in the full NETS sample except for the Revelio measures. At the industry level, the VC share is positively correlated with both trademarks and patents using the full NETS sample.

There is a positive and statistically significant relationship between customer capital intensity and knowledge capital intensity across industries, a result shown in He, Mostrom and Sufi (2026) and confirmed in the sample here. For example, the correlation between the two principal component measures of knowledge and customer capital intensity is 0.44 and statistically significantly different from 0 at the 1 percent level. This raises an interesting question: Does venture capital finance customer capital intensive industries because they are complementary to knowledge capital intensity? Or is there also a focus even on high customer capital but low knowledge capital intensity industries?

Table 3 provides the answer: venture capitalists have the largest presence in high customer capital and high knowledge capital intensive industries, but they also have high presence in high customer capital and low knowledge capital industries. Specifically, Panel A of Table 3 splits the industries into four quadrants based on whether an industry is above or below the median of the first principal component distribution on customer and knowledge capital intensity, respectively. As shown in the table, there is more mass in the high-high and low-low quadrants, which reflects the complementarity of knowledge and customer capital intensity across industries mentioned above.

The share of firms receiving a VC investment is highest in the high-high category. But it is also high in the high customer capital and low knowledge capital industries. Perhaps surprisingly, the share of VC-financed firms is lowest in the low customer capital, high knowledge capital industries. This suggests that knowledge capital intensity, on its own, is not sufficient to explain where venture capitalists invest.

Panel B of Table 3 provides more details on the underlying industries in the high customer capital half of the distribution. Specifically, for the high-high and high-low quadrants, Panel B shows the top five industries by the share financed with venture capital. Venture capitalists specialize in the high-high industries of chemical manufacturing (which includes bio-pharmaceuticals), publishing industries (which includes software companies), and high tech manufacturing (the bulk of NAICS

code 339 are medical equipment manufacturing companies). But venture capitalists also have above mean presence in high customer capital but low knowledge capital industries of household-facing manufacturing (food, beverage, apparel) and retailers.

3.2 Firm-level evidence

The strong positive relationship between customer capital intensity and VC presence holds at the firm-level as well, as shown in Table 4. The table presents regression coefficients of measures of customer and knowledge capital intensity at the firm level on whether the firm has received a VC investment in the last five years:

$$y_i = \beta \, \mathrm{VC}_i + \delta_{g(i)} + \varepsilon_i, \qquad \delta_{g(i)} \in \{\varnothing,\ \delta_{n3},\ \delta_{n3\times \mathrm{age}},\ \delta_{n3\times \mathrm{age}\times \mathrm{county}}\}, \tag{2}$$

The regression specifications begin with no fixed effects ($\delta_{g(i)}$), and then progressively add industry, industry-by age, and industry-by age-by county fixed effects.

Even after the inclusion of industry-by age-by county fixed effects, there is a positive and statistically significant relationship between whether a firm is financed by a venture capitalist and both the wage bill share of sales and marketing workers and the probability of a firm having filed for a trademark. The estimates imply that venture capital financed firms, even within an industry-age-county cell, have a 4pp higher share of the wage bill in sales and marketing and a 39pp higher likelihood of having filed for a trademark. Consistent with the existing literature, venture-capital financed firms are also more knowledge capital intensive, as measured using the wage bill share in technical occupations or the probability of filing a patent. They have a wage bill share in technical occupations that is 8pp higher and a probability of having filed for a patent that is 21pp higher.

3.3 Is venture capital special?

The NETS-based data set is linked to a number of sources that allow us to explore how venture capital differs relative to other financial intermediaries in its focus on customer and knowledge capital intensive businesses. The left panel of Figure 2 presents the mean wage bill shares in sales and marketing and technical occupations of firms financed by venture capital, private equity, private credit, banks, and finance companies. A firm is considered "financed" by a financial intermediary if it received financing within the past five years.

Firms financed by venture capital have a mean sales and marketing share that is above 35%, which is substantially higher than firms financed by any other type of financial intermediary. Private equity and private credit are higher than banks and finance companies, but none of them are within 8pp of venture capital. VC-financed firms also have a higher wage bill share in technical roles compared to other intermediaries, but the differences are less pronounced. A similar pattern holds for trademarks and patents, as shown in the right panel. VC-financed firms are more likely to have filed for a trademark and a patent relative to all other types of financial intermediaries. For patents and trademarks, the VC and PE gap and the VC and PC gap are smaller.

Overall, Figure 2 confirms the argument that venture capitalists are indeed special relative to other intermediaries, and their uniqueness runs primarily through the customer capital intensity of their portfolio companies.

Table 5 shows similar results in a multivariate framework with industry-by age-by county fixed effects. Specifically, it reports regressions from the following specification:

$$y_i = \beta_1 \mathrm{VC}_i + \beta_2 \mathrm{PE}_i + \beta_3 \mathrm{PC}_i + \beta_4 \mathrm{Bank}_i + \beta_5 \mathrm{FC}_i + \delta_{n3\times \mathrm{age}\times \mathrm{county}} + \varepsilon_i, \tag{3}$$

As column 1 shows, venture capital is the only financial intermediary associated with a higher wage bill share in sales and marketing roles. Column 2 shows that the marginal effect of venture capital on the probability of a firm having filed for a trademark is between 15 and 20pp higher than any other financial intermediary. Figure 3 shows the univariate regression coefficient for each type of financial intermediary on the four outcomes in question. For all four outcomes, venture capital is unique among all financial intermediaries.

4 Evidence from Initial Public Offerings

Firms financed with venture capital are more customer capital intensive. Is this also true of the most successful VC investments? Perhaps the most successful outcome for an early VC investment is an IPO of a portfolio firm. This section explores a sample of IPOs to test whether venture capital is positively related to customer capital intensity in this sample of winners.

4.1 IPO sample

The IPO analysis starts from Jay Ritter's IPO database, which covers U.S. initial public offerings made between 1975 and 2025. To be consistent with the analysis in the other sections of this study, we focus only on IPOs from 2010 to 2023. An important requirement is that the firm undergoing the IPO can be matched to a *gvkey* in Compustat and that the firm has observed revenue from the S-1 filing at the last pre-IPO fiscal year. This yields a starting sample of 2,184 IPOs.

A main motivation for the focus on IPOs in the analysis is that we want to examine successful exits by the investors in the firm prior to the IPO. As such, we exclude from the sample almost 900 extremely low revenue firms that IPO without having yet achieved success in the marketplace. Specifically, we exclude any IPO for a firm that has revenue of less than $25M in 2022 dollars in the year prior to the IPO. As discussed in Appendix Section E, 60% of the IPOs removed by this restriction are clinical-stage biotechnology firms that go public to fund trials before they have a product to sell. Finally, the firm must also be able to be matched with Revelio, which further excludes 276 firms. The final sample of IPOs is 1,041 firms that conduct the IPO between 2010 and 2023. More details on the sample construction are in Section E in the appendix.

A firm is classified as venture backed if PitchBook or Capital IQ records a venture round that closed strictly before the offer date, or if the Ritter database codes the offering as venture backed. If the sources disagree, then further investigation is conducted to provide a final verdict on whether the company is VC-financed. For this sample of 1,041 firms with IPOs between 2010 and 2023, 46% are VC-financed.[8]

A main advantage of the IPO sample is that firms report financials for the period just prior to the IPO through an S-1 filing. These data are recorded in Compustat and Capital IQ, and the data set used here includes them. This allows for a unique window into the financials of firms that are still private, and this section compares those financials across firms that are and are not financed by a venture capitalist prior to the IPO.

[8] As a comparison, Kaplan and Lerner (2010) find that 60% of IPOs conducted between 1999 and 2009 have VC backing. Gornall and Strebulaev (2021) find that 42% of publicly traded companies that IPO after 1978 were VC-backed. If we include the firms with revenue below $25M in 2022 dollars, the VC-financed share in our sample is 52%.

4.2 Results from the IPO sample

Even among the most successful firms in the economy which conduct an IPO, VC-financed firms have significantly higher investment in customer capital. Table 6 presents estimates from the specification:

$$y_f = \beta \, \mathrm{VC}_f + \delta_{\text{IPO year}} \, (+ \, \delta_{n3}) + \varepsilon_f, \tag{4}$$

where y_f represents an outcome for firm f and VC_f is an indicator for whether the firm received a VC investment at some point in its history. The first two columns focus on customer capital investment outcomes as measured by the ratio of sales and marketing expense to revenue ($^{I^{SM}}/_{R}$) in the year just prior to the IPO. As column 3 shows, firms that were at some point financed by a venture capitalist are 32pp more likely to report sales and marketing as an expense item in the S-1 filing that is issued before the IPO. Furthermore, conditional on reporting, $^{I^{SM}}/_{R}$ for VC-financed firms is 23pp higher, compared to a mean of 19%.

Similar results hold for R&D: VC-financed firms are 47pp more likely to report R&D and have $^{I^{RD}}/_{R}$ that is 27pp higher. VC financed firms also have significantly higher wage bill shares for both sales and marketing and technical workers. All of these results suggest that VC-financed firms that ultimately conduct an IPO are more intangible-capital focused just before the IPO. On the other hand, as shown in column 5, VC-financed firms have lower capital expenditure to revenue ratios. So while VC-financed firms invest more in knowledge and customer capital, they invest less in capital expenditures relative to the non-VC financed success stories.

Figure 4 shows the dramatic difference in the distribution of $^{I^{SM}}/_{R}$ for VC and non-VC financed firms in the year prior to the IPO. At every point of the distribution of the two samples, the gap in $^{I^{SM}}/_{R}$ is substantial. At the median, the gap is 24pp. At the 90th percentile of the distribution, the gap is almost 50pp.

In Panel B of Table 6, coefficient estimates are reported from a specification that includes industry fixed effects. All of the coefficients decline in magnitude, and the R-squared rises substantially. This is not surprising given the strong industry patterns in customer and knowledge capital investment. However, almost all of the differences between VC and non-VC financed firms remain statistically significant from zero even when comparing two firms within the same industry.

Does this heavier investment in customer and knowledge capital translate into a capital stock that is valued in the public market? Following the logic in Crouzet and Eberly (2023), variation in the enterprise value to physical capital ratio, traditionally called Tobin's Q, captures in part the

variation across firms in intangible capital intensity. As column 8 of Table 6 shows, firms that are financed by a VC have a much higher ratio of enterprise value to physical capital in the three years after the IPO. VC-financed firms invest more heavily in intangible capital prior to the IPO and have a significantly higher Tobin's Q after the IPO, which is consistent with the view that VC-financed companies build intangible capital value.

5 Dynamics of Customer Capital around a VC Investment

5.1 Research design

The dynamic research design begins by identifying a set of treatment firms that receive their first VC investment at some point between 2010 and 2023. All treatment firms are in the NETS and Revelio intersection, although it is not required that the firm is visible in NETS for every year of the treatment window. In fact, many of the firms receiving a VC investment that ultimately appear in NETS are visible in Revelio before they appear in NETS. The design does, however, require that the firm have at least one position active in Revelio in December of the treatment year, and it also must be visible in Revelio for the full treatment window of three years before the investment to three years afterward. Following Bai (2025), the design also requires any treatment firm to be less than or equal to 15 years old at the time of the investment in order to avoid unusual VC investments for established firms. The total number of firms that are in the treatment sample is 12,224.[9]

Both research designs are stacked annual firm-year panels over the event window from $[-3, +3]$, which is centered around the treated cohort's first VC round. Two control groups are constructed for the analysis. The first is the not-yet-treated control sample. Specifically, for this sample, the control sample is drawn from the same screened VC-backed population but untreated everywhere inside the stack's window. These control firms ultimately will receive a VC investment, but after the window in which the treatment firm is examined.

The second control sample is the never-treated control sample. For this sample, the candidate pool includes 2.7 million firms. These are NETS firms matched to Revelio that never receive a PitchBook venture capital or private equity deal of any kind. Every treatment firm is matched to a set of control firms that follow this criteria:

(i) Eligibility. A match candidate is admissible for treated firm i only if it has the same Revelio

[9] Appendix Section F discusses the treatment sample in more detail.

birth year, non-missing industry and state, positive birth-year headcount, and Revelio presence in every year of i's window $[t_0 - 3, t_0 + 3]$.

(ii) Cell. Within the same birth year, cells are tried finest-first: naics-4 $\times$ state, then naics-3 $\times$ state, naics-3 $\times$ census division, naics-2 $\times$ division, stopping at the first level with at least five admissible candidates.

(iii) Ranking. Within the cell, the ten nearest neighbours on $|\log \text{headcount}_{\text{birth}} - \log \text{headcount}^{i}_{\text{birth}}|$ are kept.

Once the control groups are set, the dynamic (event-study) specification is

$$y_{ise} = \sum_{k \in \{-3,-2,0,1,2,3\}} \beta_k \left\{ T_{is} \times \mathbf{1}[e = k] \right\} + \alpha_{is} + \lambda_{se} + \varepsilon_{ise}, \tag{5}$$

where i indexes firms, s stacks, and e the event year relative to the stack's first VC round ($e = 0$ is the round year); y_{ise} is the outcome; T_{is} is 1 if firm i is in stack s's treated arm and 0 if it is a control; $\mathbf{1}[e = k]$ are event-time dummies. As a result, β_k traces the treated–control gap at each event year relative to the omitted base year $e = -1$; α_{is} is a firm $\times$ stack fixed effect (absorbing each firm's level) and λ_{se} a stack $\times$ event-time fixed effect (absorbing each stack's common time path).[10]

5.2 Dynamics of customer capital around VC investment

Figure 5 presents the coefficients β_k for the share of employment in sales and marketing and technical occupations. Figure 6 presents the coefficients β_k for the probability that a firm obtains a patent or a trademark in the year in question. In both figures, the left panel presents the coefficients from the not-yet-treated design, and the right panel presents them from the never-treated design.

In both designs and across both outcomes, the most striking result is the rise in customer capital investment in the years just prior to the initial VC investment. From three years before to the year of the VC investment, treated firms boost the share of employment in sales and marketing roles by about 1.5pp relative to control firms. This is true in both research designs. In contrast the share of

[10] In terms of interpretation, we are careful to avoid language that implies causation when exploring the results on dynamics around the initial venture capital investment. The discussion below simply describes the outcomes from before through after the investment relative to the control groups. For studies that attempt to isolate the selection versus treatment issues in entrepreneurial finance, please see Sorensen (2007); Kerr, Lerner and Schoar (2014a), Bernstein, Giroud and Townsend (2016), and Howell (2017).

employment in technical roles remains flat.

There is some evidence in the not-yet-treated design that the share of employment in sales and marketing falls at the end of the treatment window. However, this is a spurious result of the fact that the control firms are likely to receive a VC investment just after the treatment window. These "control" firms are boosting sales and marketing employment in order to obtain VC financing, which makes the relative ratio for the treatment firms fall. Consistent with this argument, there is no such decline in the share of employment in sales and marketing when the control group is the never-treated firms, as shown in the right panel.

Firms receiving a VC investment also witness a sharp rise in the probability of obtaining a trademark just before investment. The sharp rise from $e = -1$ to $e = 0$ is especially large: almost 6pp. In both designs, there is a steady rise in the probability that treatment firms file for a patent relative to the control firms throughout the event window.

Figure 7 isolates the β_k coefficient plots to the $e = -3$ to $e = 0$ window to highlight the run-up in customer capital investment just prior to the VC investment. For trademarks, at the annual frequency, the results do not distinguish whether the sharp rise in trademarks from $e = -1$ to $e = 0$ happens in the months just before or just after the VC investment. Figure 8 utilizes the monthly data on trademark filings at our disposal, and it shows that the sharp rise occurs just before the investment.

Table 7 summarizes the results across both designs and all four outcomes. Specifically, it shows the outcome gaps between treatment and control just prior to the investment, and how the gaps evolve after the investment. As the top two rows of both panels show, the differences in employment shares for treated versus control firms exist prior to the VC investment itself. Firms are already heavy in sales and marketing and technical roles even before their first VC round.

In fact, the top row of Panel B can be directly compared to the coefficient estimate of 0.040 in column 4 of panel A of Table 4; the latter is the average difference in the sales and marketing wage bill share in the cross-section. This gap is close to the gap at $e = -1$ in the top panel of Panel B of Table 7. Put differently, the dynamic specification allows us to see that the entire gap in the cross-section exists even before the VC investment. This also offers a benchmark against which to judge the 1.5pp rise in sales and marketing employment share just before the investment. More than one-third of the gap present at $e = -1$ reflects a run up in anticipation of receiving the VC investment.

5.3 Scaling

While customer capital investment appears to rise significantly just before a VC investment, the VC investment is associated with a rapid scaling up of the business afterward. Figure 9 reports estimates of β_k from equation 5 where the outcome variable is total headcount of the firm from Revelio.[11] The results are consistent across both designs. Prior to the VC investment, the treated firms are growing more rapidly. However, the scaling of the business accelerates significantly after the investment. The scale effect is larger in the never-treated research design, but it is also large in the not-yet-treated design. In terms of magnitudes, the business scales up between 3 and 4 times relative to the control group.

Table 8 presents a regression of the annual first differences in headcount scaled by the total headcount as of the year before treatment ($e = -1$). The right hand side variable of the panel regression is the treatment times post indicator variable. Treated firms experience a 40 to 65pp larger growth rate in the years after the VC investment. The growth in technical and sales and marketing roles also is substantially higher for firms that receive a VC investment in the years directly afterward. In summary, while firms invest heavily in customer capital in the lead-up of a VC investment, the VC investment is associated with a substantial scaling up effect in its aftermath.

6 Why Does Venture Capital Specialize in Customer Capital?

In this section, we use data from the MSCI Private Capital Universe (formerly known as Burgiss) on investment returns of the portfolio company holdings of private funds to address our core economic question: why is venture capital specialized in financing customer capital? We show that investments in customer capital intensive industries share the return characteristics targeted by venture capitalists, requiring high risk tolerance and patience to gain exposure to power-law distributed returns. In contrast, these return characteristics are largely absent in knowledge capital-intensive industries.

[11] These scaling results post VC investment are consistent with studies using using completely different data sources such as European Orbis data (e.g. Aragoneses and Saxena (2025)) or US Census data (e.g. Akcigit et al. (2022))

6.1 Private Financing and the Returns on Investment of Intangible Capital

The MSCI database is widely regarded as the gold standard for studying private capital performance because it is sourced directly from limited partners (LPs) such as pension funds, which require detailed quarterly reportings from general partners (GPs), thereby free of survivorship bias (Harris, Jenkinson and Kaplan, 2014). This database provides key return and cash flow data for nearly 300 thousand investment holdings. For approximately one-third of these holdings, the database also provides detailed time-series cash flow data.[12] The precise timing of contributions and distributions allows us to decompose total value to paid-in capital ("TVPI") into realized (distributed to paid-in capital; "DPI") and unrealized (residual value to paid-in capital; "RVPI" = TVPI minus DPI) components at each point in time, yielding approximately 2 million holding-date observations. For a given holding-date observation, we measure DPI, TVPI, and DPI/TVPI, and winsorize each measure at the 99.99th percentile to mitigate the influence of extreme outliers but preserving the tail. We also remove erroneous observations for which DPI exceeds TVPI or DPI is negative.

For most underlying portfolio companies, the database reports firm founding age, allowing us to characterize these return measures over the firm life cycle, as well as GIC industry codes, which allow us to incorporate the industry-level intangibility measures studied earlier in the paper. The GIC classification has four levels of granularity: the first two digits identify the GIC Sector, the first four digits the GIC Industry Group, the first six digits the GIC Industry, and the full eight digits the GIC Sub-industry. To match the GIC classification to the NAICS 3-digit industry-level intangibility measures, we construct a crosswalk between GIC Sub-industries and NAICS 3-digit industries using Compustat, which reports both GIC and NAICS codes for each public firm. Specifically, for each GIC Sub-industry, we assign the NAICS 3-digit industry that occurs most frequently among firms in Compustat. We incorporate the industry-level intangibility measures developed earlier in the paper, constructed at the NAICS 3-digit level. This process yields intangibility measures available for approximately 80% of investments. We apply the same filters that we applied to the NETS-Revelio sample. Specifically, we drop Real Estate industries from our analysis as many private funds in MSCI engage in real estate investments that go beyond the corporate focus of our paper. We also drop some industries related to education services and financial services where our measurement outside of MSCI of sector-level knowledge and customer capital is imperfect.

We present empirical evidence on the distribution of various return characteristics at both the

[12] The approximate numbers are 100,000 holdings held by nearly 5,000 funds across 77,000 firms.

industry and investment levels. For industry level analyses, we additionally drop industries with fewer than 100 observations for data sensitivity reasons.

6.2 Industry-level Analyses

In this section, we show at the industry level that the return distributions of portfolio holdings financed by venture capitalists are similarly skewed to those of customer-capital-intensive holdings, but not to those of knowledge-capital-intensive holdings.

We start by showing in Figure 10 total investment share and VC penetration across four broad groups of sectors: high vs low customer capital and/or knowledge capital based on the industry medians of the respective principal component measures we've been using before. The largest volume of private capital investment and VC activity is directed towards the group intensive in both customer and knowledge capital, highlighting a potential complementarity between both types of intangibles. Similarly high levels of VC penetration are present in sectors with high customer capital intensity but low knowledge capital; however, high knowledge capital sectors with low customer capital attract the least amount of VC activity. To understand why venture capital may be particularly focused on financing customer capital, we conduct regression analyses relating investment return characteristics to VC penetration and measures of intangibility. For ease of interpretation, all variables entering the regressions are standardized to have a mean of 0 and a standard deviation of 1.

$$Y_j = \alpha + \beta \, \text{VCshare}_j + \varepsilon_j, \tag{6}$$

First, Figure 11 reports the estimated coefficients of regressing various characteristics of return distributions on the share of VC-backed holdings at the GIC Industry Group j level following specification (6). The data shows VC investments tend to carry significantly higher downside risk, with lower propensities to deliver any positive DPI at all (with more holdings reporting zero or missing DPI). The data also shows VC investments tend to have significantly heavier upper tails of returns across various measures of tail thickness in DPI and TVPI. In addition, the average DPI / TVPI ratio of VC holdings is significantly lower and the average residual value TVPI $-$ DPI (known in the industry as "RVPI") is higher. Finally, firm count-weighted TVPI and (especially) TVPI $-$ DPI ratios tend to be significantly higher, which indicates that most of the value at VC-backed firms is yet to be realized. This can be viewed as a measure of higher illiquidity requiring patience to

realize distributions out of investments. In addition, this can be also seen as a sign that venture capitalists actively allocate larger amounts of investment towards higher unrealized continuation value holdings given the particular significance of the weighted results.[13]

$$Y_j = \alpha + \beta \, \mathrm{CC}_j + \gamma \, \mathrm{KC}_j + \varepsilon_j, \tag{7}$$

Next, Figure 12 presents results by industry intangibility. Replacing the regressor with the industry-level customer capital and knowledge capital principal component measures (specification 7), we find that the return distribution characteristics of holdings in customer-capital-intensive industries are significantly closer to the skewed distribution targeted by venture capitalists than are those of holdings in knowledge-capital-intensive industries. Specifically, the industries that are more customer capital intensive tend to have the combination of higher downside risk (e.g. lower probability of paying any distribution) and higher upside potential (i.e. heavier DPI and TVPI return tails) that we showed earlier venture capitalists select on. In contrast, knowledge-capital-intensive sectors tend to exhibit less extreme downside risk and upside potential: they pay positive distributions more frequently and have thinner upper tails of returns than customer-capital-intensive industries. Customer capital intensive industries also tend to have lower DPI/TVPI ratios and higher investment-weighted TVPI - DPI residual values, indicating customer capital exhibits an active allocation of capital towards high unrealized value investments much. This return characteristic of customer capital intensive industries is a key feature of VC investments that is not shared by knowledge capital industries.

Figure 13 provides the same result using the underlying four sources of data used to estimate the principal component measures of customer and knowledge capital. All measures agree that customer capital intensive sectors tend to exhibit the thicker right tail of returns and nearly all show higher risk and the patience results. The two measures built from sources available for both private and public firms–(1) Revelio-based sales and marketing vs technical wage bills, (2) USPTO-based share of firms with trademarks vs patents–tend to show that knowledge intensive investments have weakly thinner tails than the average holding. In contrast, measures available only for public firms–(3) Compustat-based share of sales and marketing vs research and development expenditures to revenue ratios, and (4) Compustat-based share of S&M and R&D capital to total capital stock, the latter of which have a lot more zeroes–tend to find tails that are still thinner than those of customer

[13]The graph also shows that standard return metrics such as mean and variance fail to show differences between VC and non-VC holding returns despite the skewness in the data, calling for our more detailed distributional analysis.

capital intensive sectors, but weakly thicker than those found with non-Compustat measures. Figure A5 explores these differences comparing ordinal ranking of knowledge intensity to cardinal measures of the principal component, the later being our preferred measure because it preserves the spread in intensity better.

Figures A6 and A7 show that our results are broadly robust to additionally controlling for finance and operations workforce characteristics (from the Revelio data) and physical capital (e.g., capital expenditures-to-sales in the Compustat sample).

6.3 Holding-level Analyses

In this section, we present complementary micro-level evidence conducting tests at the holding level, where a holding represents an investment by a given fund in a given portfolio company tracked in the MSCI database.

Table 9 reports holding-level summary statistics. Nearly half of the holdings or investments in our sample are VC investments. The mean customer-capital principal component score is 1.3, while the mean knowledge-capital principal component score is 1.6. These principal component scores were constructed from the earlier datasets based on NETS and Revelio and standardized to have a mean of 0 and a standard deviation of 1. Hence, this indicates that the typical firm invested in by private capital funds in the MSCI database operates in industries that are more intensive in both customer capital and knowledge capital—by 1.3 and 1.6 standard deviations, respectively—than the average firm in the underlying datasets. Such a pattern is also evident in the indicators based on the industry median. According to the Revelio wage bill-based measures, 58% of investments are flagged as being in industries with high customer capital and high knowledge capital, compared with 72% according to the Compustat capital share-based measures.

The mean probability of positive DPI is 27% for firms younger than 10 years, 40% for firms younger than 20 years, and 46% for firms younger than 30 years. Similarly, the mean ratio of DPI to TVPI is 29%, 39%, and 42% for the three age thresholds, respectively. In contrast, the mean probability that TVPI exceeds the top 1% of the distribution is uniformly 1% across all three age thresholds. These patterns suggest that younger firms exhibit greater downside risk and slower value realization but not necessarily greater upside potential.

Given the importance of firm age and the tendency of venture capitalists to invest in younger firms, we conduct the analysis separately for firms below each of the three age thresholds: 10,

20, and 30 years. Our outcome variables are an indicator for positive DPI, an indicator for TVPI exceeding the top 1% of the distribution, and the ratio of DPI to TVPI. For a given holding i belonging to a firm in industry j, we regress each outcome on the principal component measures of customer capital and knowledge capital developed earlier in the paper (specification 8). We also separately regress the outcomes on an indicator for VC financing, estimated both without and with GIC subindustry fixed effects (specifications 9 and 10).

$$Y_{ij} = \alpha + \beta * CC_j + \gamma * KC_j + \epsilon_{ij} \tag{8}$$

$$Y_{ij} = \alpha + \omega * I(VC)_i + \epsilon_{ij} \tag{9}$$

$$Y_{ij} = \alpha_j + \omega * I(VC)_i + \epsilon_i \tag{10}$$

Table 10 reports the main holding-level results. The customer capital principal component is negatively associated with the probability of positive DPI andthe DPI-to-TVPI ratio, and positively associated with the probability that TVPI exceeds the top 1%. In contrast, the knowledge capital principal component exhibits either the opposite relationships or no statistically significant relationships with these outcomes. These results are consistent with the patterns documented in our earlier sectoral analyses. The VC indicator exhibits a similar pattern to the customer capital principal component: VC financing is associated with a lower probability of positive DPI, a lower DPI-to-TVPI ratio, and a higher probability of extremely high TVPI.

Including industry fixed effects attenuates the estimated VC effects on the probability of positive DPI and the DPI-to-TVPI ratio, although the coefficients remain statistically significant and retain the same sign in most cases. The only exception is the positive-DPI indicator for firms younger than 10 years, for which the coefficient becomes positive and statistically significant. In contrast, the estimated effects on the probability of TVPI exceeding the top 1% remain largely unchanged after including industry fixed effects. These results suggest that, while venture capitalists appear to select into firms characterized by greater downside risk and slower value realization, a substantial portion of the relationship is driven by differences across industries.

We also run the regressions at the firm level, with the results reported in Appendix Table A6. The results are qualitatively similar, with the effects of the knowledge-capital measure being particularly more muted.

6.4 Discussion

We estimate the characteristics of the distribution of DPI and TVPI holding returns using detailed holding-level cash flows data from MSCI by VC presence and intangibility at the industry and holding levels. The data confirms that the VC model relies on the presence of extreme upside potential at the expense of large downside risk characteristic of a heavy-tailed power-law of returns. Moreover, VC holdings also exhibit lower DPI-to-TVPI ratios, suggesting that venture capitalists select into firms with greater illiquidity requiring patience for value realization. Consistent with the thesis of this paper, customer capital intensive holdings exhibit the very same cash flow and return characteristics that venture capitalists select on. Surprisingly, the high-tech, knowledge capital intensive holdings typically associated with venture capital exhibit return characteristics much closer to those associated with tangible capital firms. Knowledge capital investments have significantly lower downside risk of not generating positive cash flows and thinner right tails relative to customer capital, providing suggestive evidence that venture capital specializes more around customer capital.

7 Conclusion

A prevailing view within the academic literature and the policy debate is that venture capital finances knowledge capital investment among young firms engaged in research and development and technological innovation (see, for example, Draghi (2024)). Many of the results in this study support this view.

This study complements this view by showing that venture capital plays a unique role among financial intermediaries in providing finance to customer capital intensive firms. Part of the venture capital focus on customer capital comes from complementarities with knowledge capital, as hinted by the fact that the share of firms financed by venture capitalists is highest in industries which invest heavily in both types of intangibles.[14] However, venture capital presence is also surprisingly high in industries with high customer capital and low knowledge capital intensity, while venture capitalists rarely target industries that exhibit high knowledge capital intensity and low customer capital intensity.

[14]This supports the view that investments into commercialization that help firms secure future rents can motivate them to do R&D, since without rents there is no incentive to innovate (e.g. Ignaszak and Sedláček (2026)).

These patterns suggest that venture capitalists may be reluctant to finance knowledge capital absent substantive commercialization efforts to go to market. The value created by customer capital investment is future rents earned from a customer base, and the results here are consistent with the view that the potential of securing such rents are a necessary condition for venture capitalists to invest in a firm. Indeed, the results here show that venture capitalists select on customer capital intensity, and that entrepreneurs boost customer capital investment before receiving an investment.

The explanation for the specialization of venture capital on customer capital is based on return characteristics across sectors and investors sourced from the portfolio holdings of private capital funds. Customer capital intensive sectors require high risk tolerance and patience to gain exposure to a heavier-tailed power law return distribution, the return characteristics precisely targeted by venture capitalists. This return profile fits customer capital investments better than investments in knowledge capital. These findings contribute to the literature arguing that the nature of power laws of different productive technologies matter for investors when choosing where to allocate capital (e.g. Gabaix (2009)). Further macroeconomic implications remain to be explored.

References

Aghion, Philippe, Martin Aragoneses, Antonin Bergeaud, and Antonio Ughi, "Venture Capital, Innovation, and Government Investment," 2026. Working paper.

Akcigit, Ufuk, Emin Dinlersoz, Jeremy Greenwood, and Veronika Penciakova, "Synergizing ventures," *Journal of Economic Dynamics and Control*, 2022, *143*, 104427.

Aragoneses, Martin and Sagar Saxena, "Industrial Policy via Venture Capital," *Available at SSRN 5316925*, 2025.

_ , Christopher Hansman, David Licher, and Ramana Nanda, "The Nature of Firm Financing and Growth," 2026.

Arellano-Bover, Jaime, Carolina Bussotti, Matteo Paradisi, and Liangjie Wu, *The Labor Demand Implications of Brand Capital: Evidence from Trademark Transactions*, RFBerlin, Rockwool Foundation Berlin-Institute for the Economy and the ..., 2026.

Arora, Abhishek and Melissa Dell, "LinkTransformer: A Unified Package for Record Linkage with Transformer Language Models," 2023.

Azinovic-Yang, Li, John D Kepler, Ava E Speros, and Christopher R Stewart, "Rising Customer Durability, Falling Business Dynamism," *Falling Business Dynamism (March 26, 2026)*, 2026.

Bai, Jessica, "Venture Capital Cycles and the Startup Labor Market," 2025.

Barnatchez, Keith, Leland D Crane, and Ryan A Decker, "An assessment of the national establishment time series (nets) database," 2017.

Belo, Frederico, Xiaoji Lin, and Maria Ana Vitorino, "Brand capital and firm value," *Review of Economic Dynamics*, 2014, *17* (1), 150–169.

_ , Yu Li, Juliana Salomao, and Maria Ana Vitorino, "Intangible Capital Around the World," 2024.

Bernstein, Shai, Arthur Korteweg, and Kevin Laws, "Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment," *The Journal of Finance*, 2017, *72* (2), 509–538.

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

Block, Joern H, Geertjan De Vries, Jan H Schumann, and Philipp Sandner, "Trademarks and venture capital valuation," *Journal of Business Venturing*, 2014, *29* (4), 525–542.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch, "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data," *Journal of monetary economics*, 2021, *121*, 1–14.

Bronnenberg, Bart J, Jean-Pierre H Dubé, and Matthew Gentzkow, "The evolution of brand preferences: Evidence from consumer migration," *American Economic Review*, 2012, *102* (6), 2472–2508.

Brown, Gregory W., Robert S. Harris, Wendy Hu, Tim Jenkinson, Steven N. Kaplan, and David T. Robinson, "Private Equity Portfolio Companies: A First Look at Burgiss Holdings Data," 2020. Available at SSRN 3532444.

Cai, Wei, Andrea Prat, and Jiehang Yu, "Measuring organizational capital," *Journal of Accounting Research*, 2026.

Chiavari, Andrea, "Customer accumulation, returns to scale, and secular trends," *Journal of Monetary Economics*, 2026, p. 103935.

Cochrane, John H., "The Risk and Return of Venture Capital," *Journal of Financial Economics*, 2005, *75* (1), 3–52.

Crane, Leland D. and Ryan A. Decker, "Research with Private Sector Business Microdata: The Case of NETSD&B," Conference Paper / Working Paper (NBER), National Bureau of Economic Research 2020. NBER Conference Paper, July 12, 2020.

Crouzet, Nicolas and Janice Eberly, "Rents and intangible capital: A q+ framework," *The Journal of Finance*, 2023, *78* (4), 1873–1916.

_ , _ , Andrea Eisfeldt, and Dimitris Papanikolaou, "Intangible capital, firm scope, and growth," Technical Report, Working Paper 2025.

Desai, Pranav, Ekaterina Gavrilova, Rui Silva, and Margarida Soares, "The Value of Trademarks," 2022. Working paper, revised January 2025.

Dinlersoz, Emin, Nathan Goldschlag, Amanda Myers, and Nikolas Zolas, "An anatomy of US firms seeking trademark registration," in "Measuring and Accounting for Innovation in the 21st Century," University of Chicago Press, 2018.

Draghi, Mario, "The Future of European Competitiveness," Report, European Commission, Brussels September 2024. Report to the European Commission, in two parts.

Einav, Liran, Ben Klopack, and Neale Mahoney, "Selling subscriptions," *American Economic Review*, 2025, *115* (5), 1650–1671.

Ewens, Michael, Alexander Gorbenko, and Arthur Korteweg, "Venture Capital Contracts," *Journal of Financial Economics*, 2022, *143* (1), 131–158.

_ **and Matthew Rhodes-Kropf**, "Is a VC Partnership Greater Than the Sum of Its Partners?," *Journal of Finance*, 2015, *70* (3), 1081–1113.

Falato, Antonio, Dalida Kadyrzhanova, Jae Sim, and Roberto Steri, "Rising intangible capital, shrinking debt capacity, and the US corporate savings glut," *The Journal of Finance*, 2022, *77* (5), 2799–2852.

Flikkema, Meindert, Carolina Castaldi, Ard-Pieter De Man, and Marcel Seip, "Trademarks' relatedness to product and service innovation: A branding strategy approach," *Research Policy*, 2019, *48* (6), 1340–1353.

Gabaix, Xavier, "Power laws in economics and finance," *Annu. Rev. Econ.*, 2009, *1* (1), 255–294.

Garfinkel, Jon A., Erik J. Mayer, Ilya A. Strebulaev, and Emmanuel Yimfor, "Alumni Networks in Venture Capital Financing," *Journal of Financial and Quantitative Analysis*, 2025, pp. 1–31.

Gompers, Paul A., "Optimal Investment, Monitoring, and the Staging of Venture Capital," *Journal of Finance*, 1995, *50* (5), 1461–1489.

_ **and Vladimir Mukharlyamov**, "Transferable Skills? Founders as Venture Capitalists," NBER Working Paper 29907, National Bureau of Economic Research April 2022.

_ **, Will Gornall, Steven N. Kaplan, and Ilya A. Strebulaev**, "How Do Venture Capitalists Make Decisions?," *Journal of Financial Economics*, 2020, *135* (1), 169–190.

Gompers, Paul and Josh Lerner, "What drives venture capital fundraising?," 1999.

Gonzalez-Uribe, Juanita, Santiago Reyes, and Xiang Yin, "The Broader Role of Venture Capital Due Diligence," *Review of Financial Studies*, 2026, *39* (7), 2018–2063.

González-Uribe, Juanita, "Exchanges of innovation resources inside venture capital portfolios," *Journal of Financial Economics*, 2020, *135* (1), 144–168.

Gornall, Will and Ilya A. Strebulaev, "Squaring Venture Capital Valuations with Reality," *Journal of Financial Economics*, 2020, *135* (1), 120–143.

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

Gourio, Francois and Leena Rudanko, "Customer capital," *Review of Economic Studies*, 2014, *81* (3), 1102–1136.

Graham, Stuart J. H., Galen Hancock, Alan C. Marco, and Amanda Fila Myers, "The USPTO Trademark Case Files Dataset: Descriptions, Lessons, and Insights," *Journal of Economics & Management Strategy*, 2013, *22* (4), 669–705.

Gutiérrez, Germán and Thomas Philippon, "Investment-less growth: An empirical investigation," Technical Report, National Bureau of Economic Research 2016.

Guzman, Jorge and Scott Stern, "The state of American entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 US States, 1988–2014," *American Economic Journal: Economic Policy*, 2020, *12* (4), 212–243.

Hall, Bronwyn H. and Josh Lerner, "The Financing of R&D and Innovation," in Bronwyn H. Hall and Nathan Rosenberg, eds., *Handbook of the Economics of Innovation*, Vol. 1, Elsevier, 2010, pp. 609–639.

Harris, Robert S., Tim Jenkinson, and Steven N. Kaplan, "Private Equity Performance: What Do We Know?," *The Journal of Finance*, 2014, *69* (5), 1851–1882.

He, Bianca, "Financing Intangibles," *SSRN Electronic Journal*, 2025.

_ , Lauren I Mostrom, and Amir Sufi, "Investing in customer capital," *Journal of Finance, forthcoming*, 2026.

Hellmann, Thomas, "The Allocation of Control Rights in Venture Capital Contracts," *The RAND Journal of Economics*, 1998, *29* (1), 57–76.

_ , Alexander Montag, and Joacim Tåg, "Tolerating Losses for Growth: J-Curves in Venture Capital Investing," IFN Working Paper No. 1500, Research Institute of Industrial Economics (IFN), Stockholm 2024. Accessed via EconStor.

_ and Manju Puri, "The Interaction between Product Market and Financing Strategy: The Role of Venture Capital," *Review of Financial Studies*, 2000, *13* (4), 959–984.

_ and _ , "Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence," *Journal of Finance*, 2002, *57* (1), 169–197.

Hoberg, Gerard and Vojislav Maksimovic, "Product life cycles in corporate finance," *The Review of Financial Studies*, 2022, *35* (9), 4249–4299.

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

Howell, Sabrina T, "Financing innovation: Evidence from R&D grants," *American economic review*, 2017, *107* (4), 1136–1164.

_ , **Josh Lerner, Ramana Nanda, and Richard R Townsend**, "How resilient is venture-backed innovation? Evidence from four decades of US patenting," Technical Report, National Bureau of Economic Research 2020.

Hubmer, Joachim and Lukas Nord, "Investment in Demand and Dynamic Competition for Customers," *Available at SSRN 5491506*, 2025.

Ignaszak, Marek and Petr Sedláček, "Customer Acquisition, Business Dynamism, and Aggregate Growth," *Review of Economic Studies*, 2026, p. rdag060.

Jackson, Blake and Ilya A. Strebulaev, "Human Capital in Venture Capital: Evidence From 100,000 Venture Capitalists," SSRN Working Paper 6995859, SSRN 2026.

Jang, Young Soo and Steven N. Kaplan, "Venture Capital Start-up Selection," Working Paper 2026.

_ , **Dasol Kim, Amir Sufi, and Xiangyu Chen**, "The Lending Technology of Direct Lenders in Private Credit," 2025. Working paper.

Kaplan, Steven N and Josh Lerner, "It ain't broke: The past, present, and future of venture capital," *Journal of Applied Corporate Finance*, 2010, *22* (2), 36–47.

Kaplan, Steven N. and Per Strömberg, "Characteristics, Contracts, and Actions: Evidence from Venture Capitalist Analyses," *The Journal of Finance*, 2004, *59* (5), 2177–2210.

_ **and Per Strömberg**, "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts," *The Review of Economic Studies*, 2003, *70* (2), 281–315.

_ , **Berk A. Sensoy, and Per Strömberg**, "Should Investors Bet on the Jockey or the Horse? Evidence from the Evolution of Firms from Early Business Plans to Public Companies," *Journal of Finance*, 2009, *64* (1), 75–115.

Kerr, William R and Ramana Nanda, "Financing innovation," *Annual review of financial economics*, 2015, *7* (1), 445–462.

_ , **Josh Lerner, and Antoinette Schoar**, "The consequences of entrepreneurial finance: Evidence from angel financings," *The Review of Financial Studies*, 2014, *27* (1), 20–55.

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

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological Innovation, Resource Allocation, and Growth," *Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

Korteweg, Arthur and Stefan Nagel, "Risk-Adjusting the Returns to Venture Capital," *The Journal of Finance*, 2016, *71* (3), 1437–1470.

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

Larkin, Yelena, "Brand perception, cash flow stability, and financial policy," *Journal of Financial Economics*, 2013, *110* (1), 232–253.

Lerner, Josh, "Venture Capitalists and the Oversight of Private Firms," *Journal of Finance*, 1995, *50* (1), 301–318.

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

Lyonnet, Victor and Léa H. Stern, "Machine Learning About Venture Capital Choices," Working Paper, Fisher College of Business, Ohio State University 2026.

McGrattan, Ellen R and Edward C Prescott, "Unmeasured Investment and the Puzzling US Boom in the 1990s," *American Economic Journal: Macroeconomics*, 2010, *2* (4), 88–123.

Midrigan, Virgiliu and Daniel Yi Xu, "Finance and misallocation: Evidence from plant-level data," *American economic review*, 2014, *104* (2), 422–458.

Ottonello, Pablo and Thomas Winberry, "Capital, ideas, and the costs of financial frictions," Technical Report, National Bureau of Economic Research 2024.

Pearce, Jeremy and Liangjie Wu, "Market concentration and aggregate productivity: The role of demand," Technical Report, Staff Reports 2025.

Peters, Ryan H and Lucian A Taylor, "Intangible capital and the investment-q relation," *Journal of Financial Economics*, 2017, *123* (2), 251–272.

Puri, Manju and Rebecca Zarutskie, "On the Life Cycle Dynamics of Venture-Capital- and Non-Venture-Capital-Financed Firms," *Journal of Finance*, 2012, *67* (6), 2247–2293.

Robb, Alicia M. and David T. Robinson, "The Capital Structure Decisions of New Firms," *The Review of Financial Studies*, January 2014, *27* (1), 153–179.

Sahlman, William A., "The Structure and Governance of Venture-Capital Organizations," *Journal of Financial Economics*, 1990, *27* (2), 473–521.

Sorensen, Morten, "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital," *The Journal of Finance*, 2007, *62* (6), 2725–2762.

United States Patent and Trademark Office, "Trademark Case Files Dataset," https://www.uspto.gov/ip-policy/economic-research/research-datasets/trademark-case-files-dataset 2024. Updated March 2024; accessed August 25, 2026.

_ , "PatentsView," https://www.uspto.gov/ip-policy/economic-research/patentsview 2026. Accessed August 25, 2026.

Vigna, Stefano Della and Ulrike Malmendier, "Paying not to go to the gym," *american economic Review*, 2006, *96* (3), 694–719.

Vries, Geertjan De, Enrico Pennings, Joern H Block, and Christian Fisch, "Trademark or patent? The effects of market concentration, customer type and venture capital financing on start-ups' initial IP applications," *Industry and Innovation*, 2017, *24* (4), 325–345.

Table 1: Summary statistics: analysis sample

	Mean	Weighted mean
Firm characteristics		
Employment (NETS, 2022)	8.4	30.9
Firm age (years)	16.6	28.5
Financing (any deal, 2018–2022)		
VC (%)	1.51	1.80
PE (%)	1.26	3.05
Private credit (%)	0.47	1.34
Bank (%)	17.79	35.71
Finance company (%)	6.11	13.47
Outcomes		
Has patent (%)	1.72	3.89
Has trademark (%)	12.03	20.96
Patents (count)	0.176	0.450
Trademarks (count)	1.247	3.445
S&M wage bill share (%)	25.3	22.0
Tech wage bill share (%)	20.7	23.9
Firms	1,271,500	

Notes: One observation per firm. "Weighted mean" weights by NETS 2022 employment; the mean is unweighted. Employment is the NETS 2022 employee count; age is years since the NETS start year. Financing indicators are any qualifying deal of the given type in 2018–2022, in percent of firms. Patent and trademark counts are cumulative granted patents / registered trademarks through 2022; the "has" rows are the corresponding 0/1 indicators in percent. Wage bill shares are the Revelio December-2022 S&M and Tech shares of the firm's wage bill.

Table 2: Industry-level intangible capital and VC prevalence

	Revelio-matched sample					Full sample
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Customer capital						
	K^{SM}/K	$I^{\mathrm{SM}}/\mathrm{Rev}$	$(WL)^{\mathrm{SM}}/WL$	Trademark share	$\mathrm{PC1}^{\mathrm{cust}}$	Trademark share
Industry VC share	3.457*	1.461*	2.196**	4.597**	33.894**	8.364**
	(1.397)	(0.666)	(0.534)	(0.569)	(8.411)	(1.138)
Industries	73	73	73	73	73	75
R-squared	0.207	0.252	0.155	0.537	0.416	0.560
Panel B: Knowledge capital						
	K^{RD}/K	$I^{\mathrm{RD}}/\mathrm{Rev}$	$(WL)^{\mathrm{tech}}/WL$	Patent share	$\mathrm{PC1}^{\mathrm{know}}$	Patent share
Industry VC share	2.975**	2.310**	0.917	1.687**	43.322**	2.429**
	(0.686)	(0.612)	(0.591)	(0.387)	(7.922)	(0.533)
Industries	73	73	73	73	73	75
R-squared	0.665	0.634	0.024	0.418	0.680	0.492

Notes: Panel A outcomes measure customer capital: the HMS Compustat industry-median S&M capital share K^{SM}/K (S&M capital as a share of total capital), the HMS industry-median S&M expense-to-revenue ratio $I^{\mathrm{SM}}/\mathrm{Rev}$, the industry mean of the firm-level S&M wage bill share $(WL)^{\mathrm{SM}}/WL$ (Revelio), the share of firms with a registered trademark, and $\mathrm{PC1}^{\mathrm{cust}}$, the first principal component of these four measures. Panel B outcomes measure knowledge capital: the R&D capital share K^{RD}/K, the R&D expense-to-revenue ratio $I^{\mathrm{RD}}/\mathrm{Rev}$, the Tech wage bill share $(WL)^{\mathrm{tech}}/WL$ (Revelio), the share of firms with a patent, and $\mathrm{PC1}^{\mathrm{know}}$, the first principal component of these four. Each principal component is extracted from the standardized measures and is itself standardized across the 73 industries (mean zero, standard deviation one). Robust standard errors in parentheses. * p<0.05, ** p<0.01.

Table 3: Industry groups by customer- and knowledge-capital intensity

Panel A: VC share across the four industry groups

Industry group	Industries	Firms	VC-backed firms	VC share (%)
High customer, high knowledge	22	416,125	11,171	2.68
High customer, low knowledge	14	332,529	4,923	1.48
Low customer, high knowledge	14	156,448	619	0.40
Low customer, low knowledge	23	366,398	2,535	0.69
All industries	73	1,271,500	19,248	1.51

Panel B: Top 5 industries by VC share within the high-customer groups

NAICS	Industry	VC share (%)	$(WL)^{\mathrm{SM}}/_{WL}$ bill share (%)	Trademark share (%)	$(WL)^{\mathrm{tech}}/_{WL}$ bill share (%)	Patent share (%)
	Average across all industries	1.7	26.5	17.0	24.0	3.4
High customer, high knowledge						
325	Chemical manufacturing	10.2	34.5	47.1	26.3	16.9
513	Publishing industries	8.0	42.3	33.2	24.2	5.1
334	Computer and electronic product mfg.	7.4	29.5	38.8	41.9	23.8
335	Electrical equipment and appliance mfg.	5.3	32.3	38.2	36.3	19.4
339	Miscellaneous manufacturing	4.7	37.5	30.0	23.3	11.3
High customer, low knowledge						
311	Food manufacturing	5.3	33.8	46.6	16.8	2.3
312	Beverage and tobacco product mfg.	3.8	33.4	45.4	14.8	0.4
315	Apparel manufacturing	2.4	47.8	44.7	15.1	3.7
455	General merchandise retailers	2.4	31.5	17.5	13.3	1.8
424	Merchant wholesalers, nondurable goods	1.9	38.6	28.4	14.2	2.2

Notes: Industries are ranked on customer-capital intensity by $\mathrm{PC1}^{\mathrm{cust}}$ and on knowledge-capital intensity by $\mathrm{PC1}^{\mathrm{know}}$; a median split on each score defines the four groups. Panel A reports, for each group, the number of industries, sample firms, VC-backed firms, and the VC share of firms. Panel B lists the five 3-digit NAICS industries with the highest VC share within the high-customer / high-knowledge group and within the high-customer / low-knowledge group, preceded by the unweighted average across all 73 sample industries. All entries in percent.

Table 4: Firm-level intangible outcomes and VC backing

	(1)	(2)	(3)	(4)
Panel A: $^{(WL)^{\mathrm{SM}}}/_{WL}$				
VC-backed	0.111**	0.074**	0.047**	0.040**
	(0.019)	(0.016)	(0.013)	(0.012)
Observations	1,271,382	1,271,382	1,271,381	1,153,506
Adj. R-squared	0.001	0.084	0.100	0.104
Panel B: Trademark				
VC-backed	0.428**	0.383**	0.407**	0.387**
	(0.020)	(0.020)	(0.020)	(0.020)
Observations	1,271,500	1,271,500	1,271,499	1,153,618
Adj. R-squared	0.026	0.072	0.083	0.086
Panel C: $^{(WL)^{\mathrm{tech}}}/_{WL}$				
VC-backed	0.098**	0.077**	0.083**	0.076**
	(0.020)	(0.008)	(0.008)	(0.007)
Observations	1,271,382	1,271,382	1,271,381	1,153,506
Adj. R-squared	0.001	0.122	0.127	0.133
Panel D: Patent				
VC-backed	0.229**	0.214**	0.219**	0.207**
	(0.022)	(0.020)	(0.021)	(0.021)
Observations	1,271,500	1,271,500	1,271,499	1,153,618
Adj. R-squared	0.046	0.091	0.096	0.115
NAICS-3 FE	No	Yes	–	–
NAICS-3 × age band FE	No	No	Yes	–
NAICS-3 × age × county FE	No	No	No	Yes

Notes: Columns add fixed effects: (1) none; (2) 3-digit NAICS; (3) NAICS-3 × age band; (4) NAICS-3 × age band × county. Wage bill shares from Revelio; trademark and patent are 0/1 indicators for any registered trademark / granted patent. Standard errors clustered by 3-digit NAICS in parentheses. * p<0.05, ** p<0.01.

Table 5: Is it uniquely VC? All financing types jointly

	$(WL)^{\mathrm{SM}}/_{WL}$ share	Trademark	$(WL)^{\mathrm{tech}}/_{WL}$ share	Patent
	(1)	(2)	(3)	(4)
VC	0.043**	0.343**	0.069**	0.194**
	(0.011)	(0.015)	(0.007)	(0.019)
PE	-0.008	0.174**	0.025**	0.058**
	(0.006)	(0.015)	(0.005)	(0.007)
Private credit	0.008	0.137**	0.014**	0.028**
	(0.012)	(0.010)	(0.005)	(0.006)
Bank	-0.014	0.070**	0.019*	0.005**
	(0.008)	(0.008)	(0.009)	(0.001)
Finance company	-0.017**	0.018**	0.014*	0.003
	(0.005)	(0.006)	(0.006)	(0.002)
Observations	1,153,506	1,153,618	1,153,506	1,153,618
Adj. R-squared	0.104	0.098	0.133	0.119
NAICS-3 × age × county FE	Yes	Yes	Yes	Yes

Notes: Firm-level regressions of each outcome on all five financing indicators jointly (VC, PE, private credit, bank, finance company), absorbing NAICS-3 × age band × county fixed effects. Standard errors clustered by 3-digit NAICS in parentheses. * p<0.05, ** p<0.01.

Table 6: Intangible intensity at the IPO and pre-IPO VC backing

	Expenses (Compustat)					Wage bill share (Revelio)		Tobin's q
	(1) R&D reported	(2) R&D/Rev	(3) S&M reported	(4) S&M/Rev	(5) CAPX/Rev	(6) $(WL)^{\text{SM}}/WL$	(7) $(WL)^{\text{tech}}/WL$	(8) Avg. q (IPO yrs 0–3)
Panel A: IPO-year FE only								
VC-backed (pre-IPO)	0.468**	0.268**	0.322**	0.231**	-0.051**	0.098**	0.131**	18.065**
	(0.026)	(0.019)	(0.027)	(0.012)	(0.014)	(0.011)	(0.014)	(2.306)
NAICS-3 FE	No	No	No	No	No	No	No	No
Observations	1041	628	1041	705	1040	842	859	924
Adj. R-squared	0.233	0.156	0.135	0.296	0.025	0.134	0.104	0.126
Panel B: + NAICS-3 FE								
VC-backed (pre-IPO)	0.223**	0.236**	0.221**	0.152**	0.002	0.030*	0.054**	6.781*
	(0.033)	(0.026)	(0.034)	(0.015)	(0.014)	(0.014)	(0.016)	(3.059)
NAICS-3 FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1032	612	1032	694	1031	833	850	916
Adj. R-squared	0.581	0.363	0.398	0.344	0.538	0.358	0.489	0.239

Notes: Columns (1) and (3) are reporting (extensive-margin) indicators: the outcome is 0/1 for a non-missing R&D-to-revenue or S&M-to-revenue value. Columns (2), (4), and (5) are the Compustat/CIQ expense ratios among reporting firms (the intensive margin); columns (6)–(7) are Revelio wage bill shares (S&M or Tech wage bill over the total wage bill). Column (8) is the firm-level mean of Tobin's q over the fiscal year of the IPO and the three following fiscal years (years with data, post-IPO valuation). IPO-year fixed effects always included. Heteroskedasticity-robust standard errors in parentheses. * p<0.05, ** p<0.01.

Table 7: Pre/post gaps between treated firms and controls

	Pre gap at -1	Post gap at $+3$	Change	% pre	% post
Panel A: not-yet-treated design					
L^{SM}/L	3.63**	2.52**	-1.11**	144	-44
	(0.59)	(0.50)	(0.36)		
L^{tech}/L	3.82**	3.24**	-0.58	118	-18
	(0.58)	(0.51)	(0.35)		
P(any patent filing), cumulative	6.61**	13.40**	6.79**	49	51
	(0.41)	(0.55)	(0.48)		
P(any trademark filing), cumulative	13.63**	21.30**	7.68**	64	36
	(0.64)	(0.77)	(0.79)		
Panel B: matched never-treated design					
L^{SM}/L	4.65**	4.44**	-0.21	105	-5
	(0.30)	(0.27)	(0.21)		
L^{tech}/L	5.14**	4.89**	-0.24	105	-5
	(0.29)	(0.27)	(0.20)		
P(any patent filing), cumulative	8.88**	16.90**	8.01**	53	47
	(0.26)	(0.33)	(0.31)		
P(any trademark filing), cumulative	20.73**	35.41**	14.67**	59	41
	(0.40)	(0.44)	(0.51)		

Notes: Level gaps between treated firms and controls, in percentage points. Share rows are read at event years -1 and $+3$; unweighted, standard errors (in parentheses) clustered by firm. Patent and trademark rows use firm $\times$ window observations, pre $=$ any filing in $[-3,-1]$, post $=$ any filing in $[0,+3]$. % pre $= 100\times$ pre/post and % post $= 100\times$ change/post; they sum to 100. *, **: $|z| > 1.96, 2.576$.

Table 8: Annual first difference of headcount

	Total	Technical	S&M
Panel A: not-yet-treated design			
Treated × post	40.518**	16.690**	11.403**
	(3.152)	(1.301)	(1.144)
Observations	182,916	182,916	182,916
Firm clusters	11,801	11,801	11,801
Panel B: matched never-treated design			
Treated × post	65.486**	23.705**	19.267**
	(2.183)	(0.885)	(0.770)
Observations	761,466	761,466	761,466
Firm clusters	104,256	104,256	104,256

Notes: The annual headcount change as a percent of the firm's own $e = -1$ headcount, with a single pooled treated × post coefficient per column. Firm-stack and stack × event-time fixed effects, unweighted, standard errors clustered by firm. *, **: $|z| > 1.96, 2.576$.

Table 9: MSCI Holding-level Summary Statistics

	N	mean	p50	sd
CC Share (Revelio)	79000	0.311	0.338	0.105
KC Share (Revelio)	79000	0.256	0.241	0.092
I(CC High) (Revelio)	79000	0.668	1.000	0.471
I(KC High) (Revelio)	79000	0.774	1.000	0.418
CC Share (Compustat)	81000	1.992	1.400	1.835
KC Share (Compustat)	81000	1.063	0.958	1.062
I(CC High) (Compustat)	81000	0.871	1.000	0.335
I(KC High) (Compustat)	81000	0.737	1.000	0.440
I(CC Low KC High) (Revelio)	79000	0.196	0.000	0.397
I(CC High KC Low) (Revelio)	79000	0.090	0.000	0.286
I(CC High KC High) (Revelio)	79000	0.578	1.000	0.494
I(CC Low KC High) (Compustat)	81000	0.020	0.000	0.140
I(CC High KC Low) (Compustat)	81000	0.154	0.000	0.361
I(CC High KC High) (Compustat)	81000	0.717	1.000	0.451
Customer-capital PC1 score (std.: mean 0, SD 1)	79000	1.304	0.746	1.449
Knowledge-capital PC1 score (std.: mean 0, SD 1)	79000	1.611	1.493	1.903
I(VC)	98000	0.477	0.000	0.499
I(DPI Positive by Age 10)	98000	0.265	0.000	0.441
I(DPI Positive by Age 20)	98000	0.401	0.000	0.490
I(DPI Positive by Age 30)	98000	0.464	0.000	0.499
I(TVPI $>$ p(99) by Age 10)	67000	0.008	0.000	0.092
I(TVPI $>$ p(99) by Age 20)	79000	0.008	0.000	0.088
I(TVPI $>$ p(99) by Age 30)	86000	0.008	0.000	0.087
DPI/TVPI by Age 10	66000	0.290	0.000	0.433
DPI/TVPI by Age 20	79000	0.392	0.000	0.469
DPI/TVPI by Age 30	86000	0.419	0.023	0.474

Notes: This table reports the summary statistics of key variables used in the regressions for the MSCI holding-level sample. The number of observations (N) is rounded to the nearest 1,000.

Table 10: MSCI Holding-level Regression Results

Panel A: I(DPI Positive)

	Up to Age 10			Up to Age 20			Up to Age 30		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CC PC1 score	-0.034***			-0.045***			-0.056***		
	(0.008)			(0.008)			(0.007)		
KC PC1 score	0.009**			0.013**			0.011**		
	(0.004)			(0.005)			(0.006)		
I(VC)		-0.071***	0.032***		-0.101***	-0.036**		-0.182***	-0.136***
		(0.013)	(0.010)		(0.021)	(0.017)		(0.023)	(0.019)
Industry FE	No	No	Yes	No	No	Yes	No	No	Yes
N	79000	98000	98000	79000	98000	98000	79000	98000	98000
R-squared	0.01	0.01	0.12	0.02	0.01	0.09	0.02	0.03	0.09

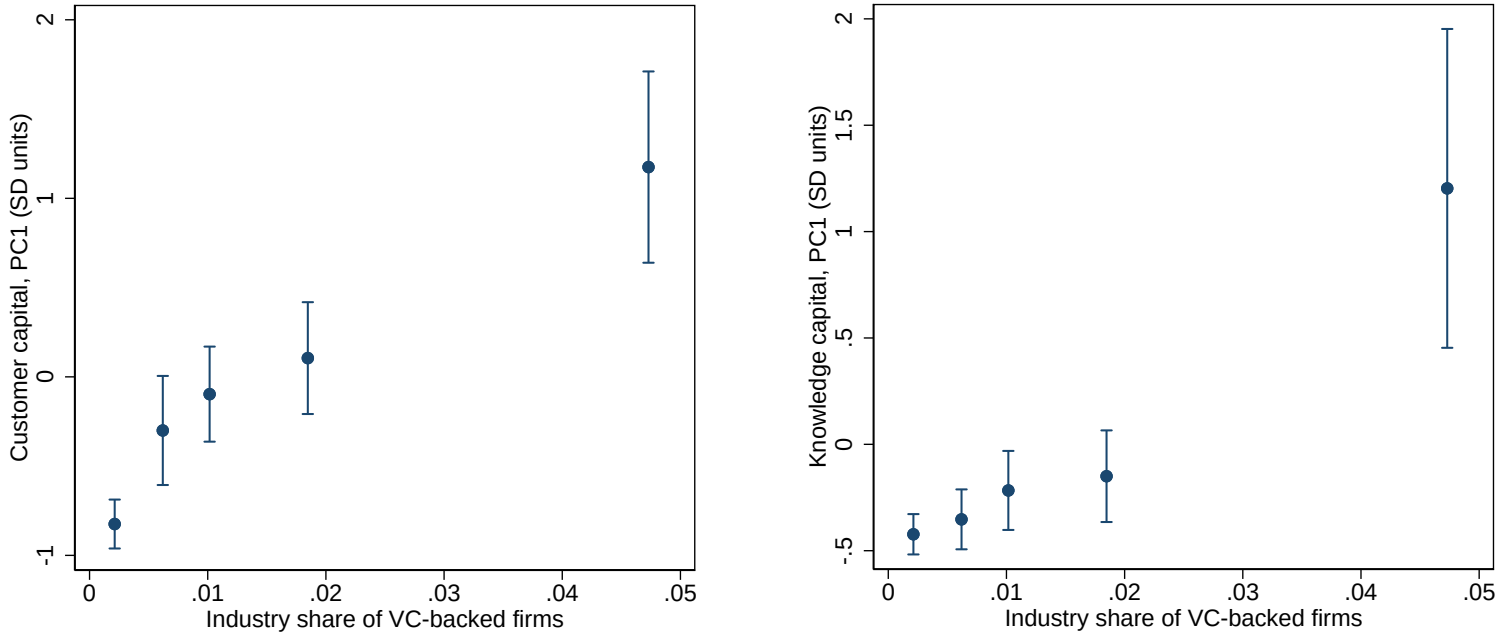
Panel B: DPI/TVPI

	Up to Age 10			Up to Age 20			Up to Age 30		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CC PC1 score	-0.050***			-0.058***			-0.060***		
	(0.008)			(0.008)			(0.008)		
KC PC1 score	0.008			0.016**			0.017**		
	(0.006)			(0.007)			(0.007)		
I(VC)		-0.213***	-0.101***		-0.163***	-0.093***		-0.169***	-0.116***
		(0.016)	(0.013)		(0.022)	(0.018)		(0.022)	(0.019)
Industry FE	No	No	Yes	No	No	Yes	No	No	Yes
N	53000	66000	66000	64000	79000	79000	70000	86000	86000
R-squared	0.03	0.06	0.12	0.03	0.03	0.09	0.03	0.03	0.09

Panel C: I(TVPI $>$ p(99))

	Up to Age 10			Up to Age 20			Up to Age 30		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CC PC1 score	0.002***			0.002***			0.002***		
	(0.000)			(0.000)			(0.000)		
KC PC1 score	-0.001***			-0.001***			-0.001**		
	(0.000)			(0.000)			(0.000)		
I(VC)		0.009**	0.008**		0.010**	0.009***		0.010**	0.010***
		(0.005)	(0.004)		(0.004)	(0.003)		(0.004)	(0.004)
Industry FE	No	No	Yes	No	No	Yes	No	No	Yes
N	53000	67000	67000	64000	79000	79000	70000	86000	86000
R-squared	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01

Notes: This table reports the holding-level estimation results from regressing key performance metrics on measures of industry intangibility and a VC financing indicator. Panel A reports results where the outcome variable is an indicator for whether the investment delivers a final positive DPI. Panel B reports results where the outcome variable is the investment holding's final DPI/TVPI. Panel C reports results where the outcome variable is an indicator for whether the investment delivers a final TVPI above the top 1%. Columns 1–3 show results for firms up to age 10, columns 4–6 for firms up to age 20, and columns 7–9 for firms up to age 30. Columns 1, 4, and 7 use the customer capital and knowledge capital principal components as regressors, with standard errors clustered at the GIC subindustry level. Columns 2, 5, and 8 use an indicator for VC financing as the regressor. Columns 3, 6, and 9 additionally include GIC subindustry fixed effects, with standard errors clustered at the fund level. The number of observations (N) is rounded to the nearest 1,000. * p<0.10, ** p<0.05, *** p<0.01.

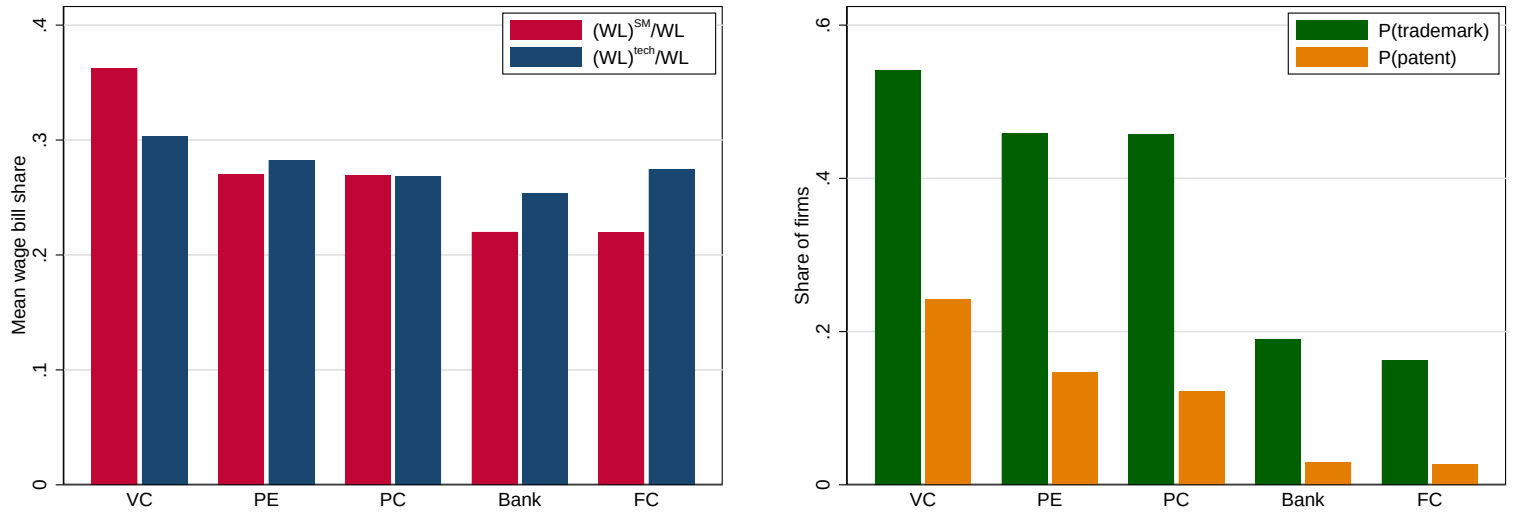


(a) Customer capital: standardized $\mathrm{PC1}^{\mathrm{cust}}$

(b) Knowledge capital: standardized $\mathrm{PC1}^{\mathrm{know}}$

Figure 1: Industry customer- and knowledge-capital composites vs. industry VC share: binned scatters at the 3-digit NAICS level

Notes: Five equal-count bins with 90% confidence intervals, computed on all 73 sample industries. The vertical axis is the first principal component of the four customer-capital measures (panel a) or of the four knowledge-capital measures (panel b), standardized across industries to mean zero and standard deviation one.

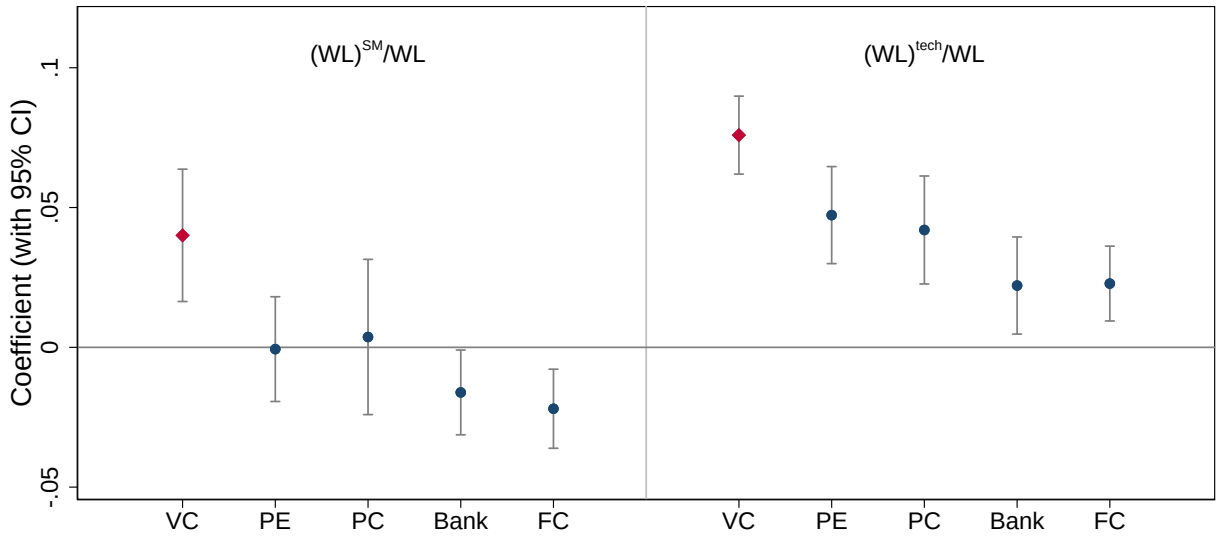


(a) Wage bill shares

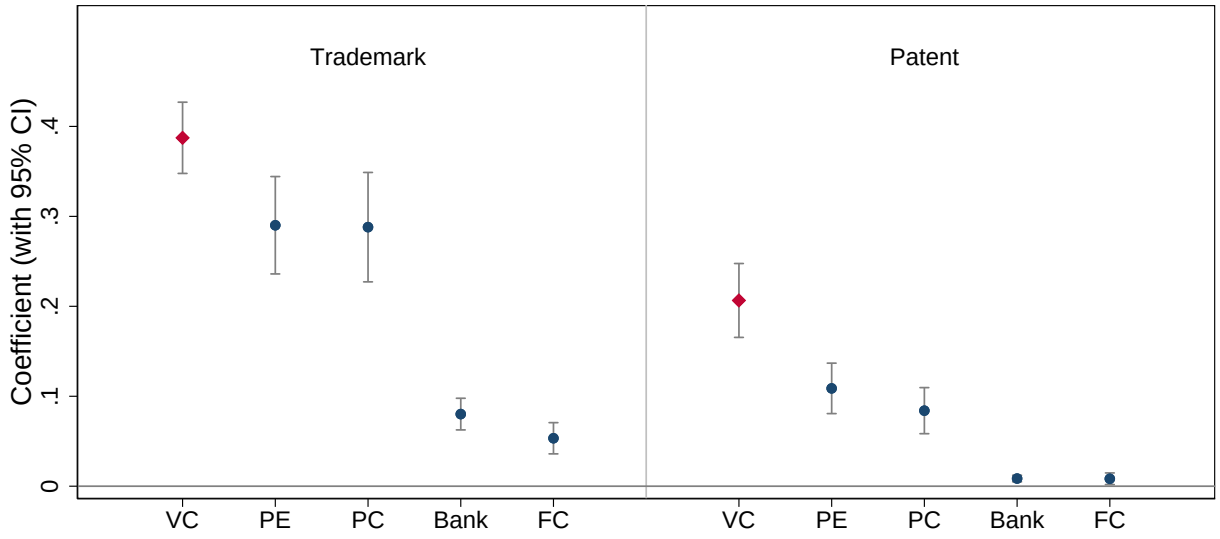
(b) Trademarks and patents

Figure 2: Raw portfolio means by financing type

Notes: Unconditional means over each financing type's portfolio companies (VC, PE, private credit, bank, finance company). Panel (a): mean S&M and Tech wage bill shares. Panel (b): shares of firms with any registered trademark and any granted patent.



(a) Wage bill shares



(b) Trademarks and patents

Figure 3: Coefficient on each financing indicator, separate firm-level regressions

Notes: Each point is the coefficient on the indicated financing indicator from a separate regression, absorbing NAICS-3 × age band × county fixed effects; bars are 95% confidence intervals with standard errors clustered by 3-digit NAICS. PC = private credit, FC = finance company.

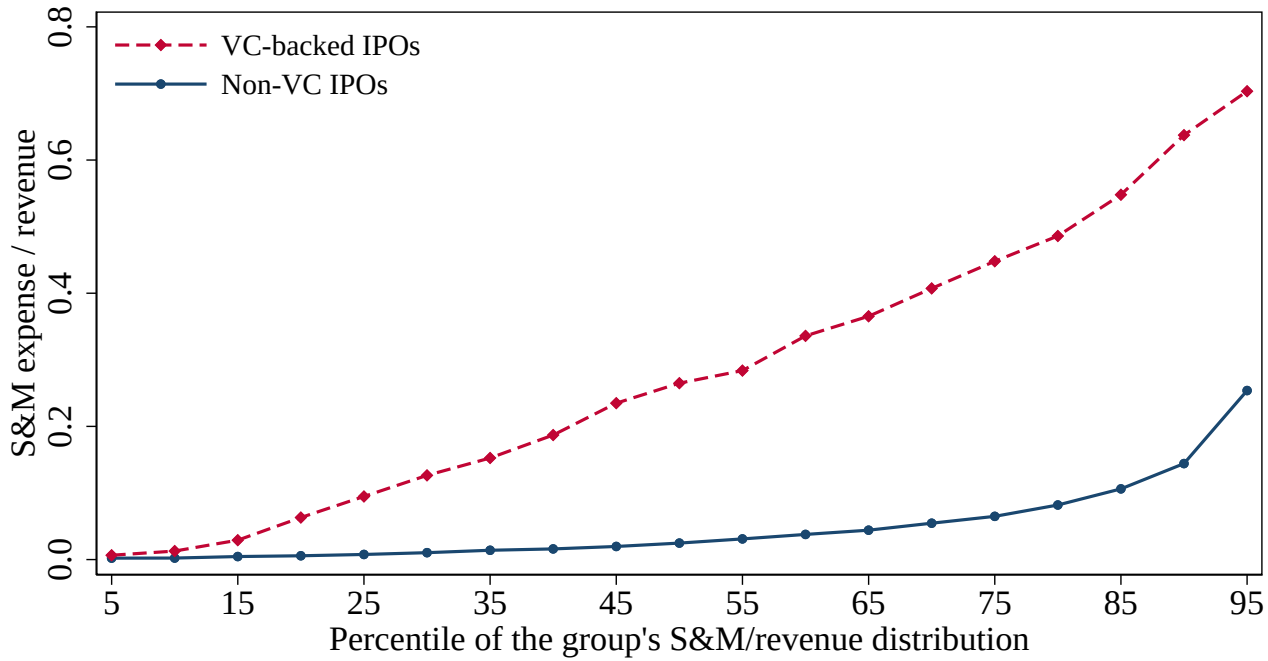
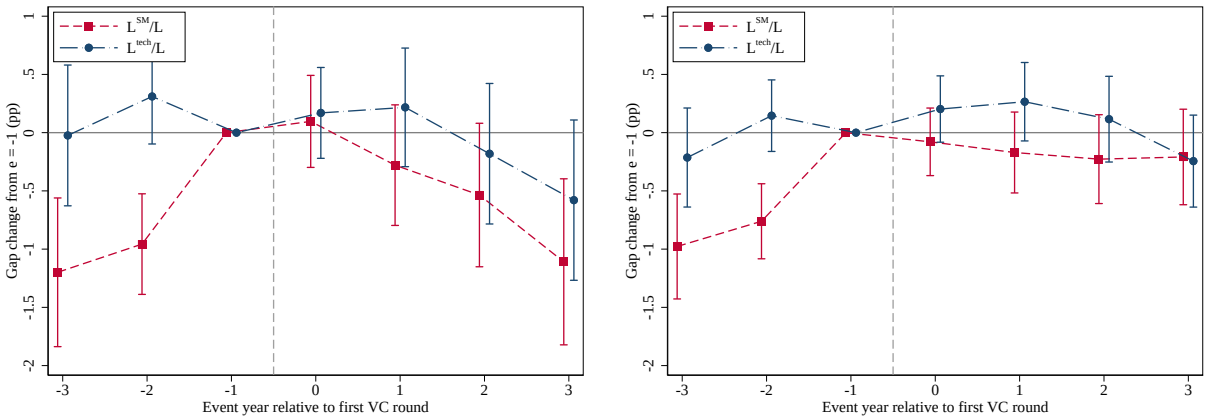


Figure 4: Distribution of S&M expense over revenue: VC-backed vs. non-VC IPOs

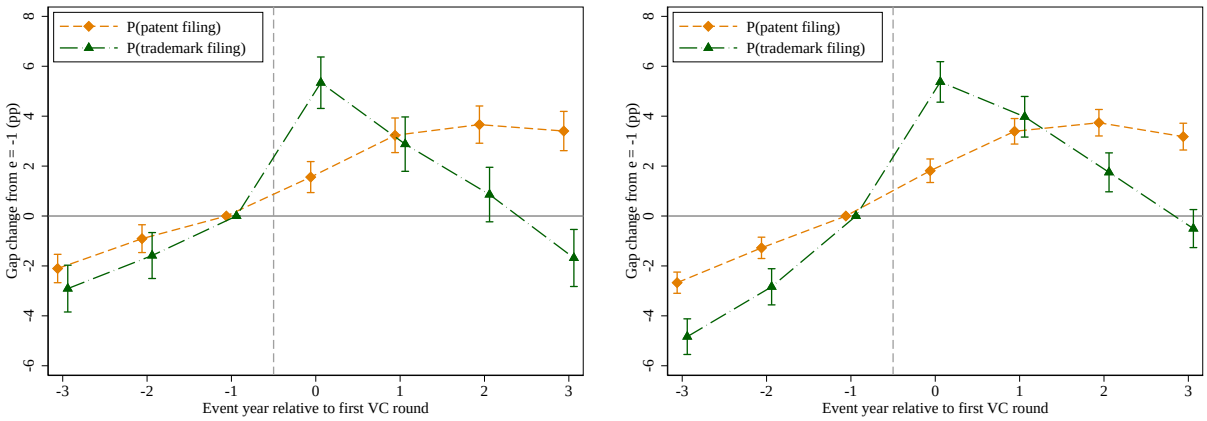
Notes: Each line plots the S&M expense over revenue ratio at the 5th through 95th percentile, in steps of five, of its own group's distribution, measured at the last pre-IPO fiscal year.



(a) Not-yet-treated design (b) Never-treated design

Figure 5: Employment shares around the first VC round: event studies

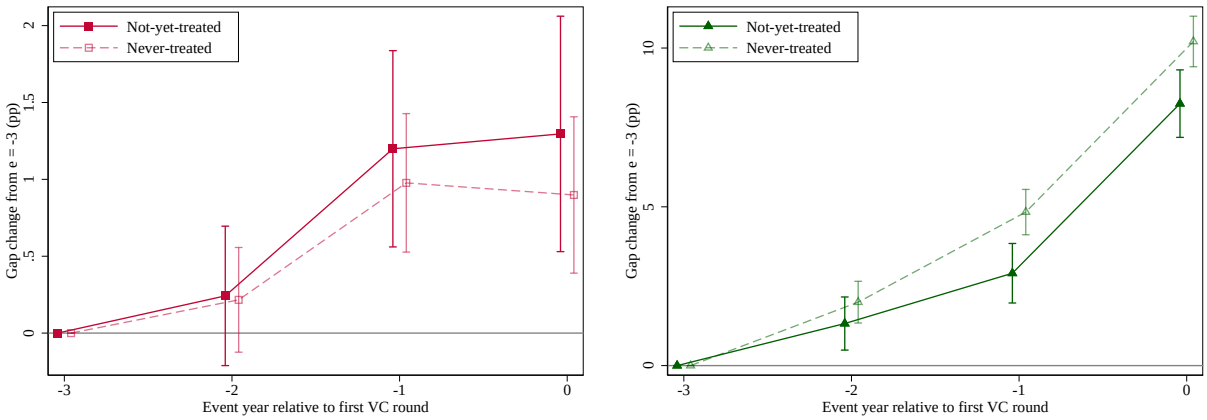
Notes: S&M and technical shares of headcount, base period $e = -1$, in percentage points with 95% bars, standard errors clustered by firm. Left: controls are firms not yet treated as of the cohort's $t + 3$. Right: matched never-treated controls.



(a) Not-yet-treated design (b) Never-treated design

Figure 6: Trademark and patent filing around the first VC round: event studies

Notes: Linear probability models for the probability of any trademark filing and any patent filing in the year, base period $e = -1$, in percentage points with 95% bars, standard errors clustered by firm. Designs as in Figure 5.



(a) S&M employment share

(b) P(trademark filing)

Figure 7: Run-up into the round, $[-3, 0]$, both designs: customer capital

Notes: Estimated on the $[-3, 0]$ window with $e = -3$ as the omitted base, percentage points, 95% bars, standard errors clustered by firm. Solid/filled: not-yet-treated design; dashed/hollow: matched never-treated design.

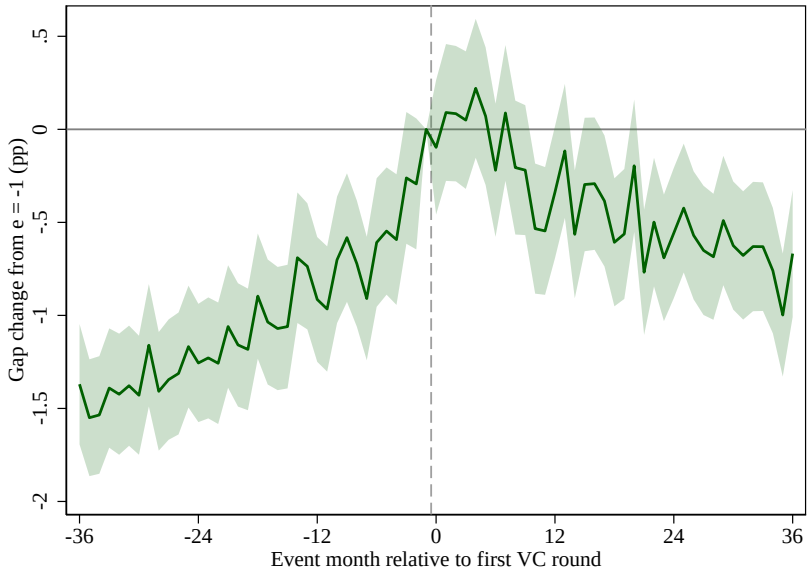


Figure 8: Monthly trademark event study, matched never-treated design

Notes: Probability of any trademark filing in the month (linear probability model), event months $[-36, +36]$ around the first VC round, base $m = -1$, matched never-treated design. Shaded band is the 95% confidence interval, standard errors clustered by firm.

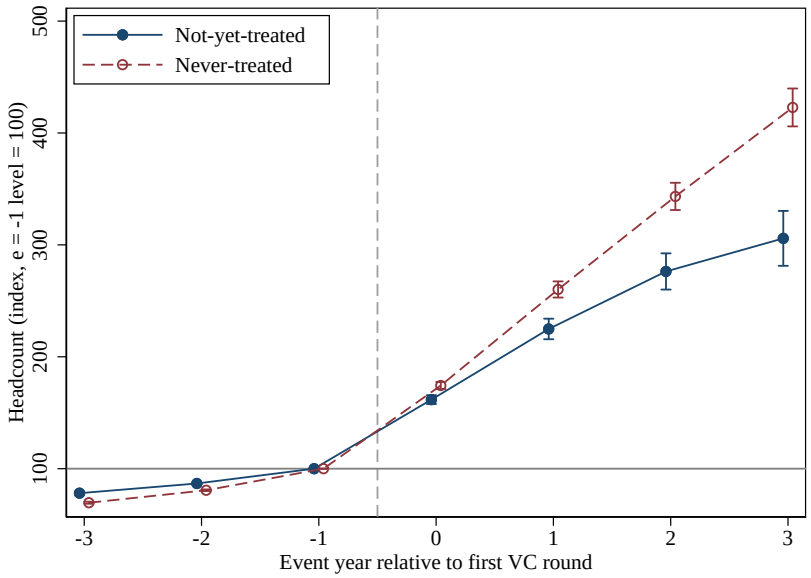


Figure 9: The scale effect: headcount around the first VC round, both designs

Notes: Total headcount scaled by the firm's own $e = -1$ headcount, re-indexed so the base period $e = -1$ equals 100; 95% bars, standard errors clustered by firm. Solid/filled: not-yet-treated design; dashed/hollow: matched never-treated design. Points are treated–control event-study gaps added to the base, not the treated firms' raw index.

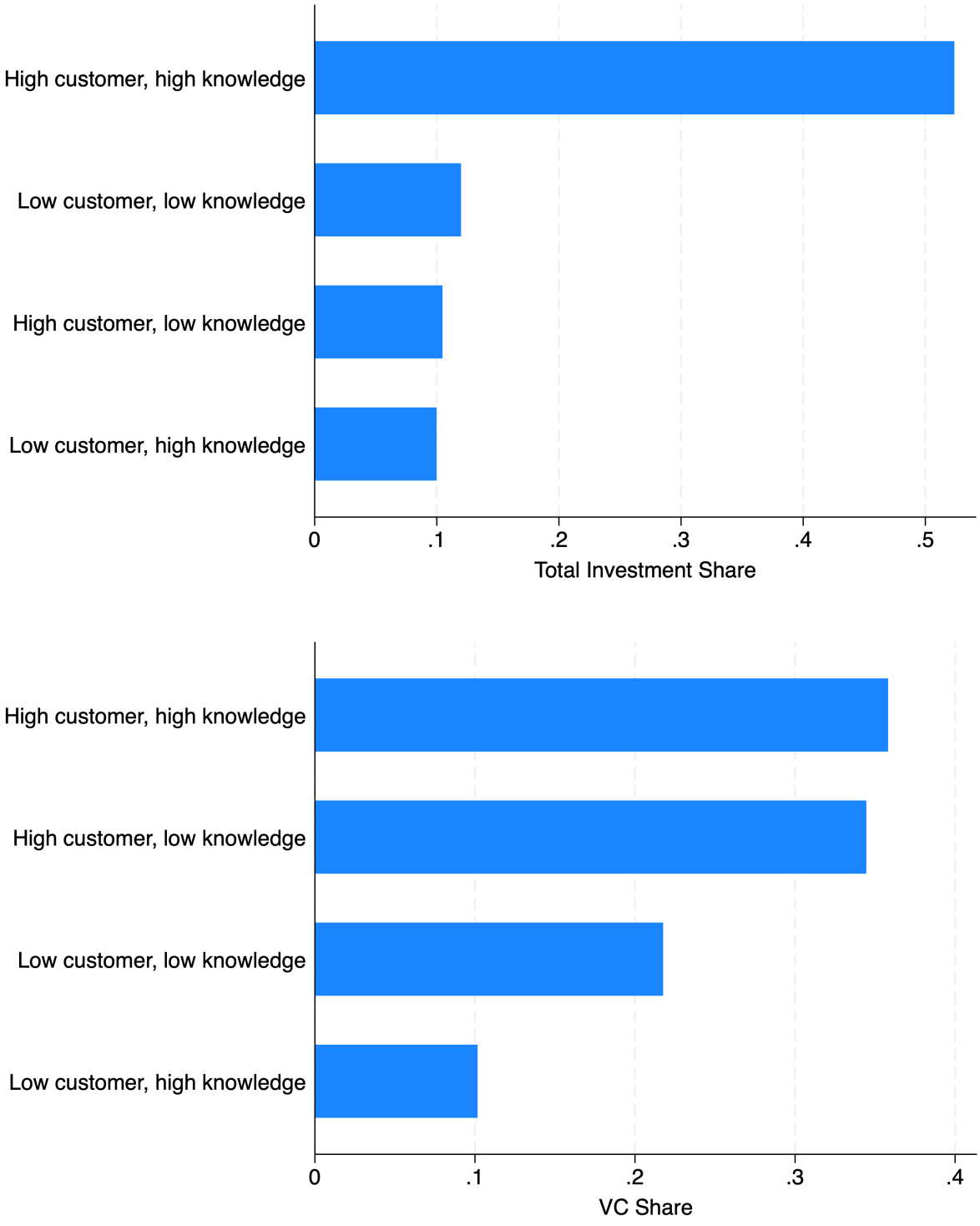


Figure 10: Venture capital, customer capital, and knowledge capital in MSCI

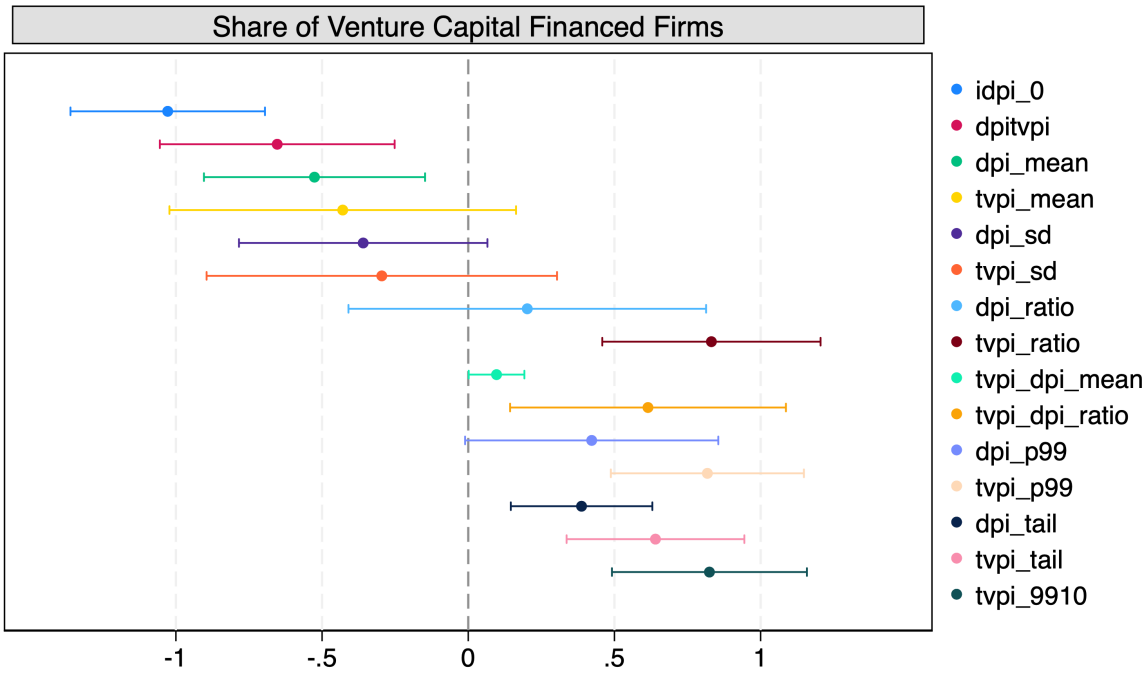


Figure 11: The Skewed Return Distribution of Portfolio Holdings Financed by VCs in MSCI

Notes: This figure reports the regression coefficients from Eq. (1), with the share of venture capital financed firms in the industry in MSCI on the right-hand side and various return characteristics on the left-hand side. These variables, measured at the holding-by-time level in the MSCI data, are aggregated to the GICS industry group level. The return variables include: (1) $idpi_0$ share with positive DPI; (2) $dpiovertvpi_{mean}$ mean DPI / TVPI; (3) dpi_{mean}, (4) $tvpi_{mean}$, (5) dpi_{sd}, (6) $tvpi_{sd}$ mean / standard deviation of DPI and TVPI; (7) dpi_{ratio} and (8) $tvpi_{ratio}$ are the investment-weighted average DPI and TVPI, (9) $tvpi-dpi_{mean}$ and (10) $tvpi-dpi_{ratio}$ is the unweighted and investment-weighted means of undistributed TVPI net of DPI, and the extremity of the upper tail for each performance measure: (11) dpi_{p99} and (12) $tvpi_{p99}$ top 1% of DPI; and TVPI; (13) dpi_{tail}, (14) $tvpi_{tail}$ top 1% to the top 10%, (15) $tvpi_{9910}$ top 1% minus bottom 10%. 95% Confidence Intervals shown. Regressions are weighted by the count of entities (portfolio firms) in each sector bin. Heteroskedasticity-robust standard errors are used.

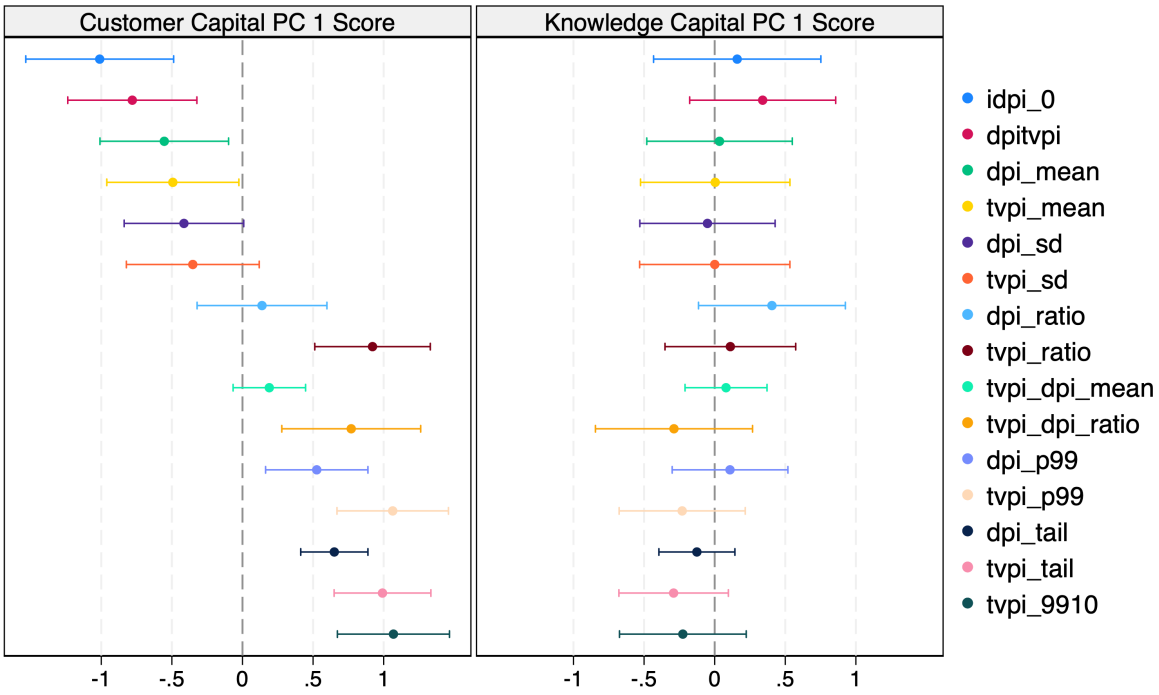
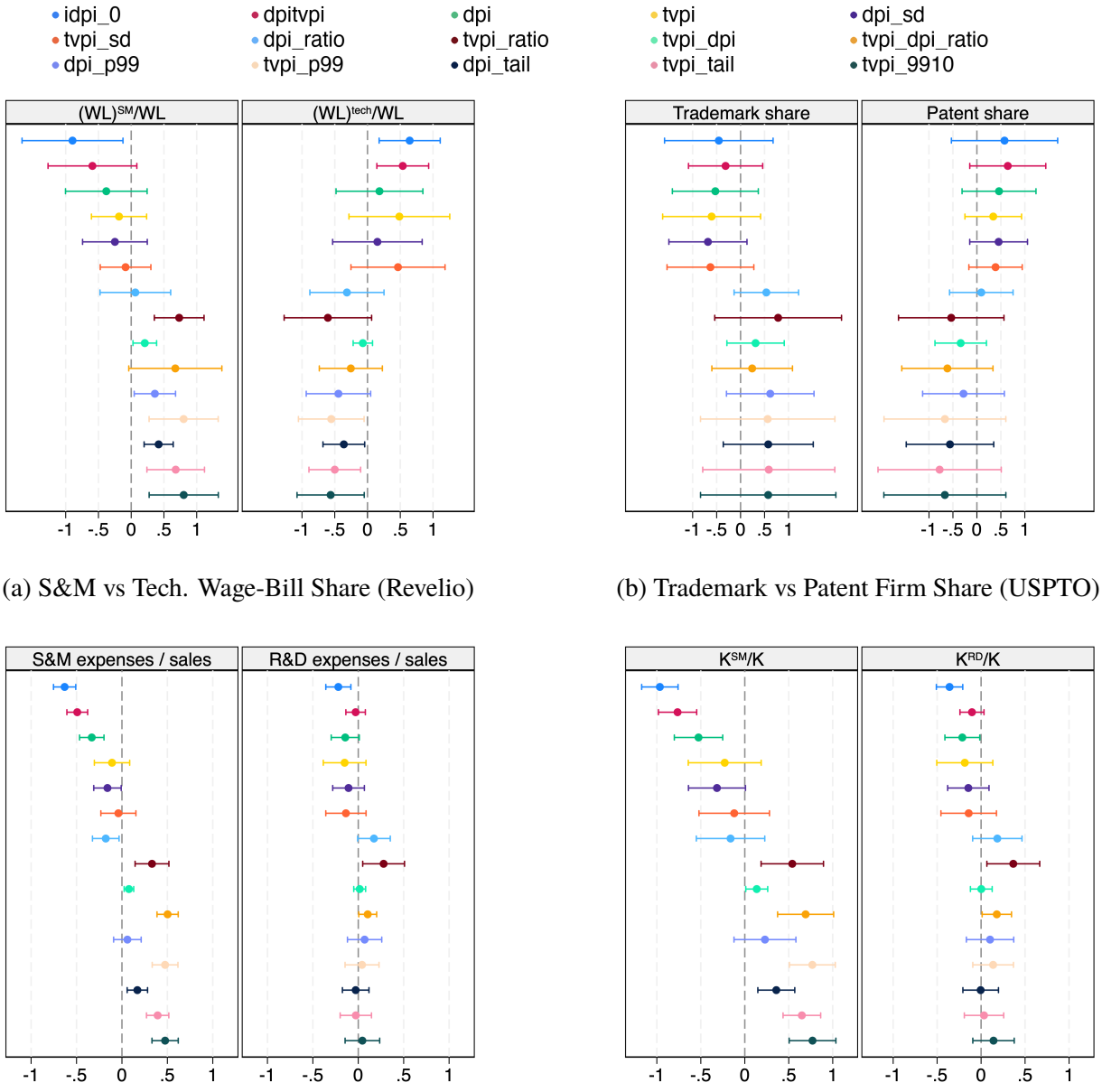


Figure 12: The Return Characteristics of Customer vs Knowledge Intensive Industries in MSCI

Notes: This figure reports the regression coefficients from Eq. (1), with the Principal Component (PC)-based measure of industry intangibility on the right-hand side and various return characteristics on the left-hand side. The two intangibility measures of the industry are included in the multi-variate regression at the same time: customer capital intensity and knowledge capital intensity. The return variables, measured at the holding-by-time level in the MSCI data, are aggregated to the GICS industry group level. The return variables include: (1) $idpi_0$ share with positive DPI; (2) $dpiovertvpi_{mean}$ mean DPI / TVPI; (3) dpi_{mean}, (4) $tvpi_{mean}$, (5) dpi_{sd}, (6) $tvpi_{sd}$ mean / standard deviation of DPI and TVPI; (7) dpi_{ratio} and (8) $tvpi_{ratio}$ are the investment-weighted average DPI and TVPI, (9) $tvpi - dpi_{mean}$ and (10) $tvpi - dpi_{ratio}$ is the unweighted and investment-weighted means of undistributed TVPI net of DPI, and the extremity of the upper tail for each performance measure: (11) dpi_{p99} and (12) $tvpi_{p99}$ top 1% of DPI; and TVPI; (13) dpi_{tail}, (14) $tvpi_{tail}$ top 1% to the top 10%, (15) $tvpi_{9910}$ top 1% minus bottom 10%. 95% Confidence Intervals shown. Regressions are weighted by the count of entities (portfolio firms) in each sector bin. Heteroskedasticity-robust standard errors are used.



(a) S&M vs Tech. Wage-Bill Share (Revelio)

(b) Trademark vs Patent Firm Share (USPTO)

(c) S&M vs R&D Investment Share (Compustat)

(d) S&M vs R&D Capital Share (Compustat)

Figure 13: Return characteristics and industry intangibility in MSCI

Notes: This figure reports Eq. (1) coefficients for RHS variables that are Revelio-based (a) S&M and technical wage bill shares, USPTO-based (b) share of firms with trademarks vs patents, Compustat-based (c) S&M vs R&D expenditures to revenues, and Compustat-based (d) S&M and R&D capital share of total capital. See description of return LHS outcomes in Figure 12. 95% Confidence Intervals shown. Regressions are weighted by the count of entities (portfolio firms) in each sector bin. Heteroskedasticity-robust standard errors are used.

Internet Appendix

Table A1: VC-financed firms as a share of NETS and of Census (SUSB) firm counts, by employment size bin, 2022

Employees	VC-financed firms	NETS firms	VC share of NETS (%)	Census (SUSB) firms	VC share of Census (%)	NETS/ Census
Panel A: Raw NETS employment						
<5	32,309	25,041,636	0.13	4,029,041	0.80	6.2
5–9	7,623	2,406,476	0.32	1,034,135	0.74	2.3
10–19	4,610	730,866	0.63	656,917	0.70	1.1
20–99	4,594	421,862	1.09	561,160	0.82	0.8
All <100	49,136	28,600,840	0.17	6,281,253	0.78	4.6
Panel B: Payroll employment concept (one owner subtracted per firm)						
<5	23,871	14,493,230	0.16	4,029,041	0.59	3.6
5–9	6,454	1,965,207	0.33	1,034,135	0.62	1.9
10–19	3,821	527,865	0.72	656,917	0.58	0.8
20–99	4,188	385,299	1.09	561,160	0.75	0.7
All <100	38,334	17,371,601	0.22	6,281,253	0.61	2.8

Notes: VC-financed firms and NETS totals are the full NETS–PB/CIQ sample, binned on firm-level NETS employment; Census counts are 2022 SUSB employer firms, U.S. totals across all industries, by enterprise employment size. The VC share of Census places the NETS-based VC count over the SUSB denominator. NETS counts nonemployer businesses and imputes positive employment to most of them, so the NETS denominator overstates the number of small employer businesses; Panel B subtracts one (non-payroll owner) employee per firm and drops firms at zero, an employer-universe approximation.

Table A2: VC-backed firms in the 2022 cross-section: size, age, geography, and timing

	Full sample		Revelio analysis sample		
	Firms	% of col.	Firms	% of col.	% in analysis
Panel A: by firm size in 2022 (employees)					
1	10,877	22.1	2,979	15.5	27.4
2	12,551	25.5	4,006	20.8	31.9
[3,5]	11,320	23.0	4,776	24.8	42.2
[6,10]	6,454	13.1	2,914	15.1	45.2
[11,20]	3,821	7.8	2,077	10.8	54.4
[21,50]	3,065	6.2	1,808	9.4	59.0
[51,100]	1,123	2.3	688	3.6	61.3
Panel B: by firm age in 2022 (years)					
1	4,905	10.0	1,490	7.7	30.4
2	5,443	11.1	1,887	9.8	34.7
3	6,885	14.0	2,411	12.5	35.0
4	6,408	13.0	2,414	12.5	37.7
5	4,881	9.9	1,963	10.2	40.2
[6,9]	11,343	23.0	4,962	25.8	43.7
[10,15]	5,822	11.8	2,603	13.5	44.7
[16,29]	2,643	5.4	1,185	6.2	44.8
30+	881	1.8	333	1.7	37.8
Panel C: by county (top 20, headquarters)					
36061 New York (Manhattan), NY	2,961	6.0	1,333	6.9	45.0
06075 San Francisco, CA	2,936	6.0	1,433	7.4	48.8
06037 Los Angeles, CA	2,886	5.9	1,069	5.6	37.0
06085 Santa Clara, CA	1,963	4.0	968	5.0	49.3
25017 Middlesex (Cambridge), MA	1,395	2.8	598	3.1	42.9
48453 Travis (Austin), TX	1,186	2.4	441	2.3	37.2
53033 King (Seattle), WA	1,130	2.3	529	2.7	46.8
06081 San Mateo, CA	1,077	2.2	500	2.6	46.4
06073 San Diego, CA	1,056	2.1	520	2.7	49.2
17031 Cook (Chicago), IL	988	2.0	401	2.1	40.6
06001 Alameda (Oakland), CA	981	2.0	450	2.3	45.9
06059 Orange, CA	789	1.6	334	1.7	42.3
25025 Suffolk (Boston), MA	776	1.6	334	1.7	43.0
48201 Harris (Houston), TX	769	1.6	281	1.5	36.5
36047 Kings (Brooklyn), NY	746	1.5	254	1.3	34.0
04013 Maricopa (Phoenix), AZ	690	1.4	205	1.1	29.7
12086 Miami-Dade, FL	628	1.3	186	1.0	29.6
48113 Dallas, TX	572	1.2	195	1.0	34.1
13121 Fulton (Atlanta), GA	505	1.0	214	1.1	42.4
08031 Denver, CO	478	1.0	178	0.9	37.2
All other counties	24,336	49.5	8,774	45.6	36.1
County unassigned	363	0.7	51	0.3	14.0
Panel D: by year of the latest VC round, 2018–2022					
2018	6,068	12.3	1,953	10.1	32.2
2019	6,723	13.7	2,402	12.5	35.7
2020	6,991	14.2	2,711	14.1	38.8
2021	13,047	26.5	5,215	27.1	40.0
2022	16,382	33.3	6,967	36.2	42.5
All VC-backed firms	49,211	100.0	19,248	100.0	39.1

Notes: One row per bin, VC-backed firms only. "Full sample" = the 2022 NETS–PB/CIQ universe conditioned on 1–100 employees; "Revelio analysis sample" = the subset in the paper's analysis sample. "% of col." is the column's distribution across the panel's bins; "% in analysis" = the share of the bin's full-sample firms in the analysis sample. Panel B: age as of 2022. Panel C sorts counties by the full-sample count. Panel D: latest round year in the window.

Table A3: Predictive regressions of the Revelio match

	(1)	(2)	(3)	(4)	(5)
2 employees	0.83**			0.93**	0.81**
	(0.17)			(0.15)	(0.25)
3–5 employees	4.08**			3.45**	3.52**
	(0.50)			(0.44)	(0.73)
6–10 employees	7.79**			6.75**	8.00**
	(1.97)			(1.73)	(1.24)
11–20 employees	20.55**			17.73**	18.45**
	(2.66)			(2.49)	(2.17)
21–50 employees	29.04**			25.15**	25.56**
	(3.08)			(2.97)	(2.62)
51–100 employees	36.36**			31.41**	31.64**
	(2.84)			(2.80)	(2.46)
Age 3–5		1.48**		1.47**	1.23**
		(0.39)		(0.37)	(0.40)
Age 6–9		2.37**		1.73**	1.36*
		(0.59)		(0.51)	(0.52)
Age 10–15		3.10**		1.78**	1.21*
		(0.57)		(0.56)	(0.56)
Age 16–29		5.66**		2.80**	2.21**
		(0.65)		(0.57)	(0.60)
Age 30+		9.29**		3.25**	2.59**
		(0.73)		(0.61)	(0.50)
VC			29.80**	27.78**	26.38**
			(2.30)	(1.74)	(1.35)
PE			26.91**	20.29**	19.18**
			(1.39)	(1.02)	(0.83)
Private credit			7.15**	3.67**	3.16**
			(0.72)	(0.39)	(0.41)
Bank			11.25**	7.01**	6.98**
			(1.55)	(1.20)	(1.17)
Finance company			1.86	0.69	1.28
			(0.98)	(0.69)	(0.66)
NAICS-3 FE					✓
Mean dep. var. (%)	4.45	4.45	4.45	4.45	4.45
Observations	28,605,288	28,605,288	28,605,288	28,605,288	28,605,288
Adj. R-sq.	0.053	0.014	0.023	0.066	0.075

Notes: Linear probability models on the full sample; the dependent variable is 100× (matched to Revelio), so coefficients are percentage points and the sample mean is 4.45. Omitted categories: one employee, age 1–2. Financing indicators are any qualifying deal of the given type in 2018–2022. Standard errors clustered by 3-digit NAICS in parentheses; * $p < 0.05$, ** $p < 0.01$.

Table A4: Firm-level trademark and patent outcomes: full NETS–PB/CIQ sample

	(1)	(2)	(3)	(4)
Panel A: Trademark				
VC-backed	0.348**	0.330**	0.335**	0.318**
	(0.041)	(0.040)	(0.040)	(0.039)
Observations	28,605,288	28,605,288	28,605,288	28,364,248
Adj. R-squared	0.009	0.025	0.032	0.041
Panel B: Patent				
VC-backed	0.151**	0.147**	0.148**	0.142**
	(0.026)	(0.026)	(0.026)	(0.025)
Observations	28,605,288	28,605,288	28,605,288	28,364,248
Adj. R-squared	0.018	0.030	0.034	0.054
NAICS-3 FE	No	Yes	–	–
NAICS-3 × age band FE	No	No	Yes	–
NAICS-3 × age × county FE	No	No	No	Yes

Notes: Sample-selection robustness: Table 4's trademark and patent panels re-estimated on the full NETS–PB/CIQ sample, without requiring the Revelio match. Columns as in Table 4. Standard errors clustered by 3-digit NAICS in parentheses. * $p<0.05$, ** $p<0.01$.

Table A5: Is it uniquely VC? Trademark and patent horserace, full NETS–PB/CIQ sample

	Trademark	Patent
	(1)	(2)
VC	0.288**	0.134**
	(0.033)	(0.024)
PE	0.172**	0.050**
	(0.019)	(0.008)
Private credit	0.087**	0.015**
	(0.015)	(0.004)
Bank	0.038**	0.003**
	(0.007)	(0.001)
Finance company	0.004	0.001
	(0.003)	(0.000)
Observations	28,364,248	28,364,248
Adj. R-squared	0.046	0.056
NAICS-3 × age × county FE	Yes	Yes

Notes: Sample-selection robustness: Table 5's trademark and patent columns re-estimated on the full NETS–PB/CIQ sample, with the same fixed effects. Standard errors clustered by 3-digit NAICS in parentheses. * p<0.05, ** p<0.01.

Table A6: MSCI Firm-level Regression Results (Robustness)

Panel A: I(DPI Positive)

	Age 10			Age 20			Age 30		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CC PC1 score	-0.032***			-0.044***			-0.055***		
	(0.010)			(0.008)			(0.007)		
KC PC1 score	0.005			0.007			0.005		
	(0.004)			(0.006)			(0.006)		
I(VC)		-0.085***	0.044***		-0.112***	-0.026		-0.186***	-0.123***
		(0.015)	(0.011)		(0.021)	(0.016)		(0.022)	(0.018)
Industry FE	No	No	Yes	No	No	Yes	No	No	Yes
N	50000	65000	65000	50000	65000	65000	50000	65000	65000
R-squared	0.01	0.01	0.12	0.02	0.01	0.09	0.02	0.03	0.09

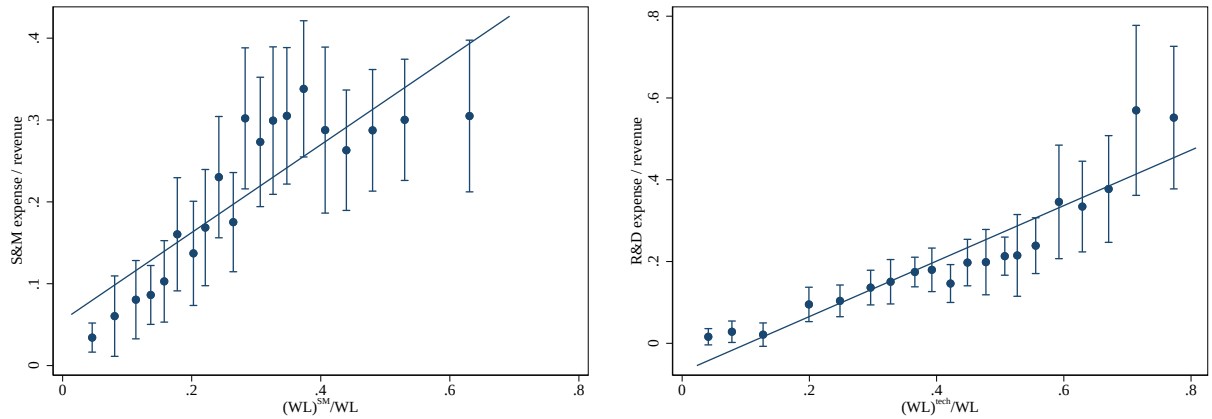
Panel B: DPI/TVPI

	Age 10			Age 20			Age 30		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CC PC1 score	-0.046***			-0.053***			-0.056***		
	(0.010)			(0.009)			(0.009)		
KC PC1 score	0.001			0.009			0.010		
	(0.007)			(0.008)			(0.008)		
I(VC)		-0.227***	-0.098***		-0.172***	-0.082***		-0.175***	-0.105***
		(0.016)	(0.012)		(0.020)	(0.016)		(0.020)	(0.016)
Industry FE	No	No	Yes	No	No	Yes	No	No	Yes
N	35000	46000	46000	40000	53000	53000	44000	57000	57000
R-squared	0.03	0.07	0.13	0.03	0.03	0.10	0.03	0.04	0.09

Panel C: I(TVPI $>$ p(99))

	Age 10			Age 20			Age 30		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CC PC1 score	0.003***			0.003***			0.003***		
	(0.000)			(0.000)			(0.000)		
KC PC1 score	-0.001**			-0.001*			-0.000		
	(0.000)			(0.000)			(0.000)		
I(VC)		0.014***	0.012***		0.014***	0.013***		0.015***	0.013***
		(0.003)	(0.002)		(0.002)	(0.002)		(0.002)	(0.002)
Industry FE	No	No	Yes	No	No	Yes	No	No	Yes
N	35000	46000	46000	41000	53000	53000	44000	57000	57000
R-squared	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.01

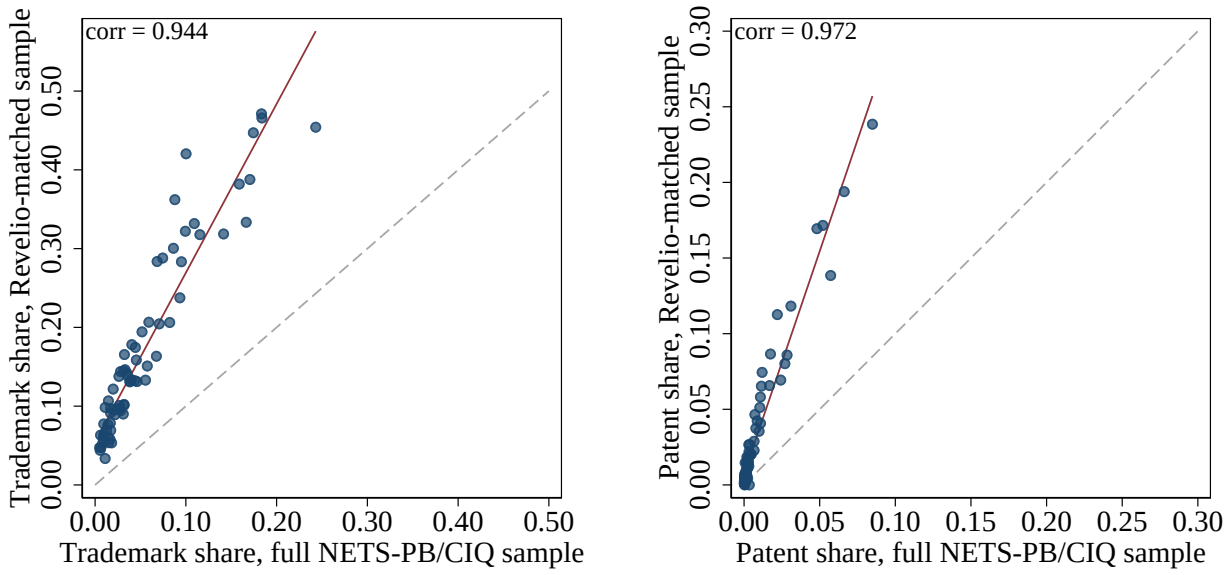
Notes: This table reports the firm-level estimation results from regressing key performance metrics on measures of industry intangibility and a VC financing indicator. Panel A reports results where the outcome variable is an indicator for whether the investment delivers a final positive DPI. Panel B reports results where the outcome variable is the investment's final DPI/TVPI. Panel C reports results where the outcome variable is an indicator for whether the investment delivers a final TVPI above the top 1%. Columns 1–3 show results for firms up to age 10, columns 4–6 for firms up to age 20, and columns 7–9 for firms up to age 30. Columns 1, 4, and 7 use the customer capital and knowledge capital principal components as regressors, with standard errors clustered at the GIC subindustry level. Columns 2, 5, and 8 use an indicator for VC financing as the regressor. Columns 3, 6, and 9 additionally include GIC subindustry fixed effects, with standard errors clustered at the fund level. The number of observations (N) is rounded to the nearest 1,000. * p<0.10, ** p<0.05, *** p<0.01.



(a) S&M expense vs. $(WL)^{\mathrm{SM}}/WL$ (b) R&D expense vs. $(WL)^{\mathrm{tech}}/WL$

Figure A1: Accounting vs. workforce measures of intangible intensity at the IPO

Notes: Binned scatters (20 bins) with linear fit, IPO-year fixed effects absorbed, on the common IPO sample at the last pre-IPO year. Raw correlations: 0.38 between S&M expense/revenue and the S&M wage bill share; 0.46 between R&D/revenue and the Tech wage bill share.



(a) Trademark share

(b) Patent share

Figure A2: Industry IP shares: full NETS–PB/CIQ sample vs. Revelio-matched subsample

Notes: Sample-selection robustness for the industry IP measures. Each point is a 3-digit NAICS industry; the horizontal axis is the share of the industry's firms with a registered trademark (panel a) or granted patent (panel b) computed on the full NETS–PB/CIQ sample, the vertical axis is the same share computed on the Revelio-matched subsample. Dashed line: 45 degrees; solid line: OLS fit.

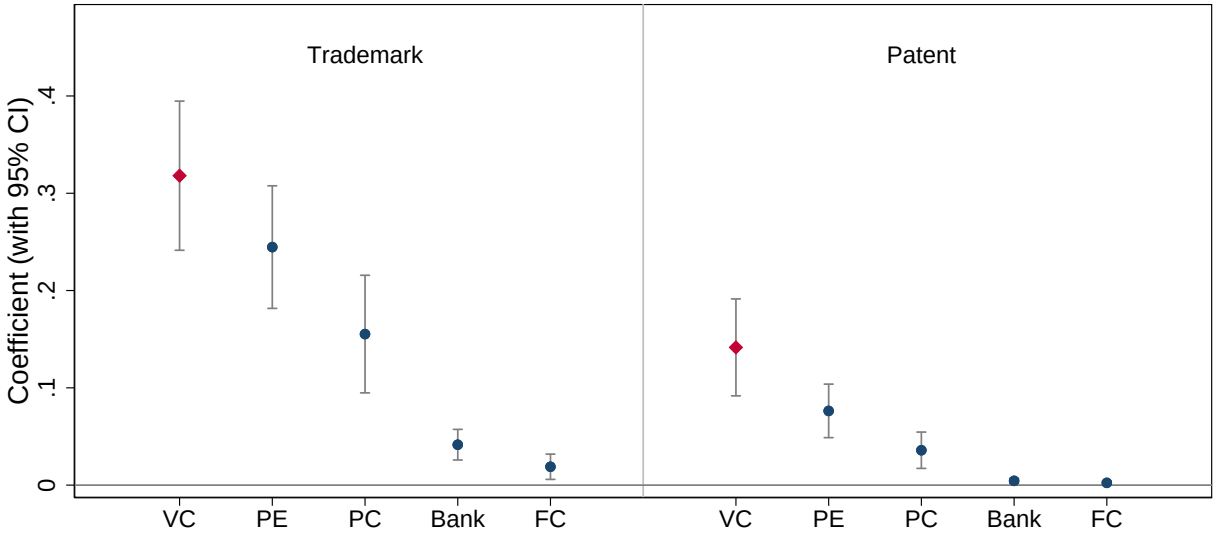
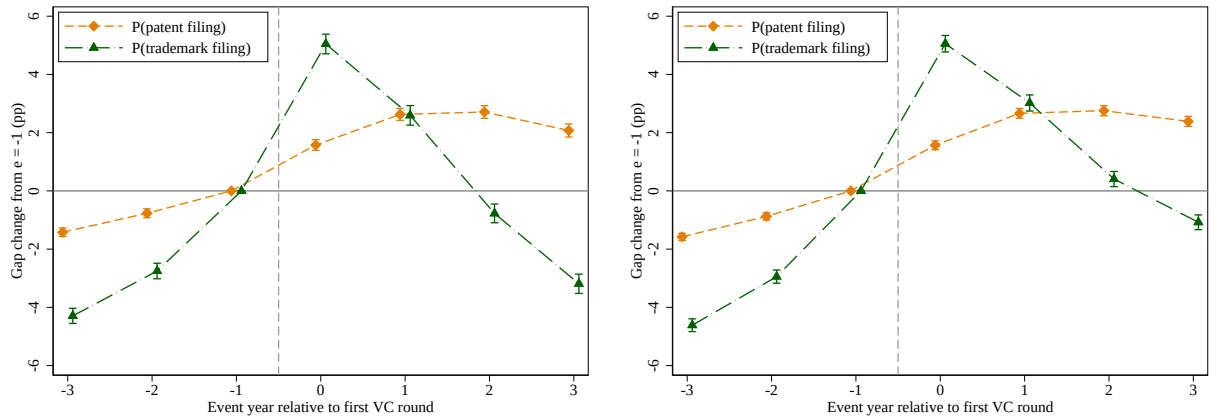


Figure A3: Coefficient on each financing indicator, trademarks and patents: full NETS–PB/CIQ sample

Notes: Sample-selection robustness: Figure 3(b) re-estimated on the full NETS–PB/CIQ sample. Each point is the coefficient on the indicated financing indicator from a separate single-indicator regression absorbing NAICS-3 $\times$ age band $\times$ county fixed effects; bars are 95% confidence intervals with standard errors clustered by 3-digit NAICS. PC = private credit, FC = finance company.

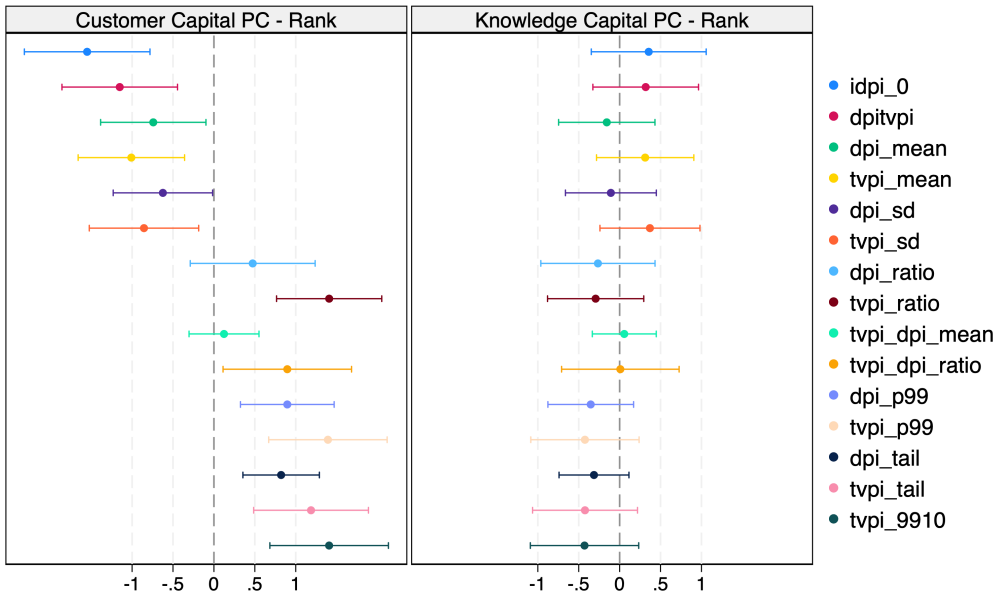


(a) Not-yet-treated design

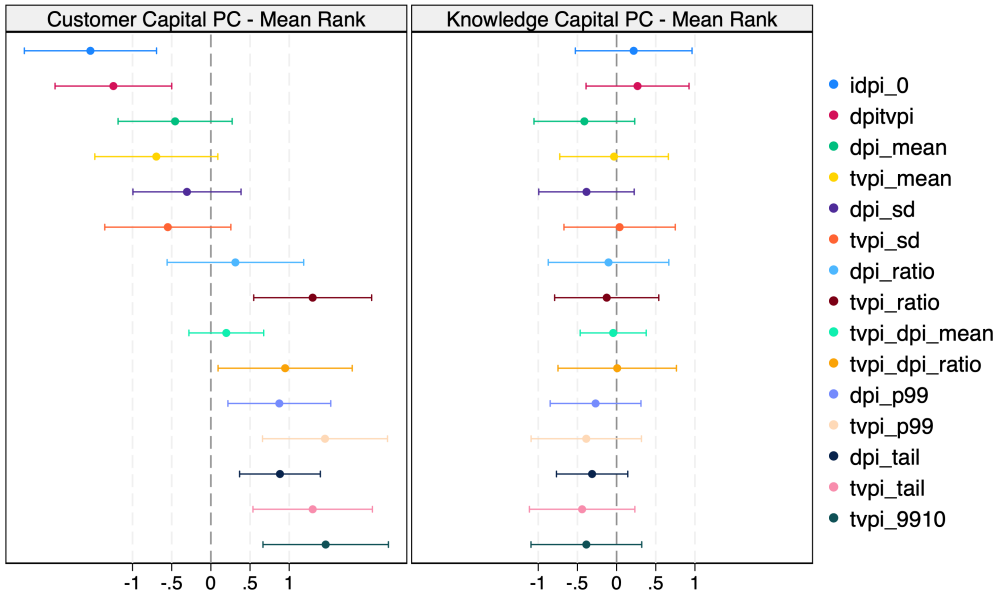
(b) Never-treated design (NETS-matched)

Figure A4: Trademark and patent filing around the first VC round: full NETS–PB/CIQ sample

Notes: Sample-selection robustness for Figure 6: the same event studies re-estimated on the full treated sample, with no Revelio presence or balance requirement. Panel (a): not-yet-treated design. Panel (b): never-treated design with NETS-native matching in place of the Revelio rule. Specifications unchanged: base $e = -1$, percentage points, 95% bars, standard errors clustered by firm.



(a) Principal Component (-) Overall Ranking



(b) Principal Component (-) Average of 4 Sub-Rankings

Figure A5: MSCI Returns & Industry Intangibility: Principal Component Ranking Measures

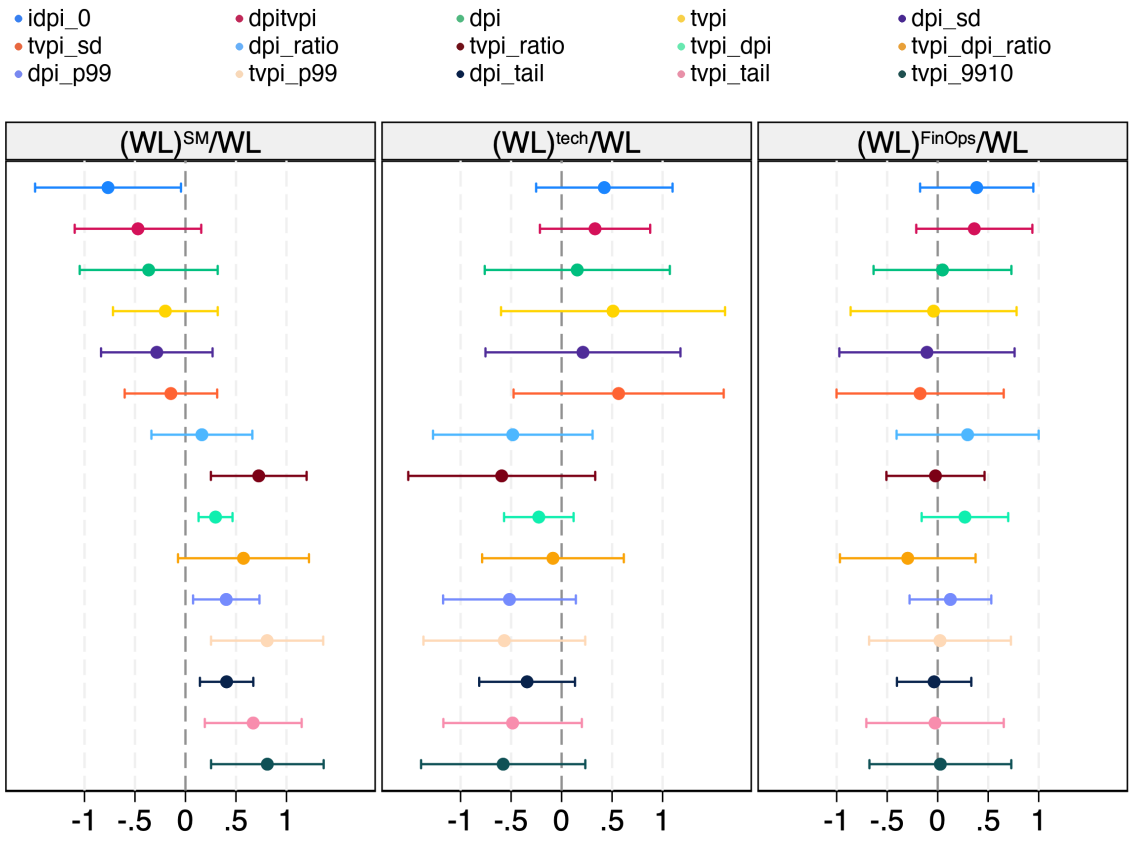
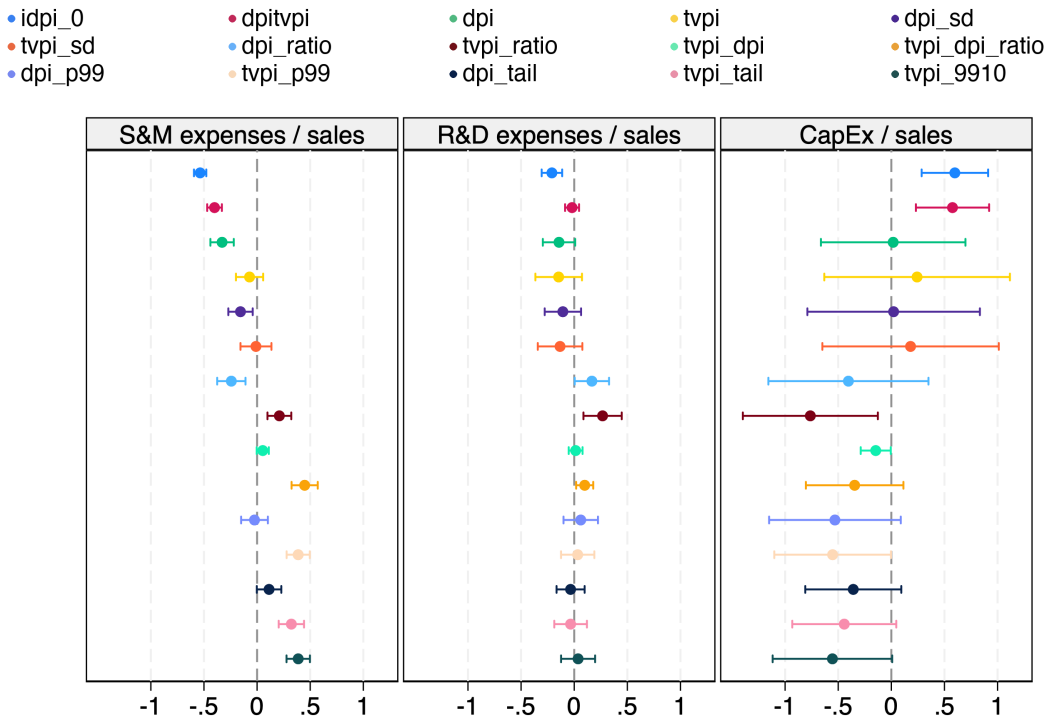
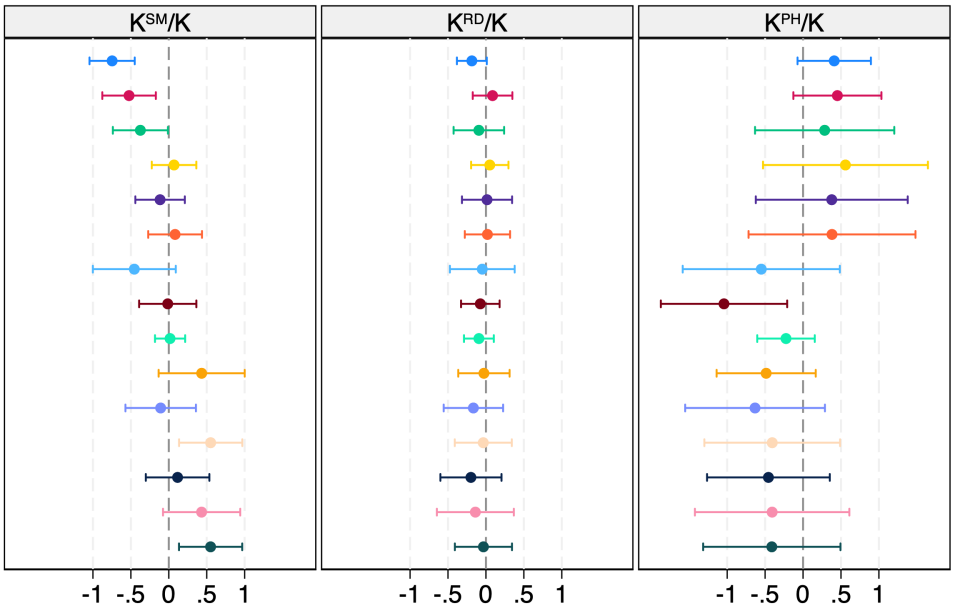


Figure A6: Return characteristics and industry tangibility (Revelio)

Notes: This figure reports the regression coefficients from Eq. (1), with the Revelio-based industry intangibility measures on the right-hand side and various return characteristics on the left-hand side. Three tangibility measures of the industry are included in the multi-variate regression at the same time: customer capital intensity, knowledge capital intensity, as well as finance and operations intensity. The return variables, measured at the holding-by-time level in the MSCI data, are aggregated to the GICS industry group level. The return variables include: (1) $idpi_0$ share with positive DPI; (2) $dpiovertvpi_{mean}$ mean DPI / TVPI; (3) dpi_{mean}, (4) $tvpi_{mean}$, (5) dpi_{sd}, (6) $tvpi_{sd}$ mean / standard deviation of DPI and TVPI; (7) dpi_{ratio} and (8) $tvpi_{ratio}$ are the investment-weighted average DPI and TVPI, (9) $tvpi - dpi_{mean}$ and (10) $tvpi - dpi_{ratio}$ is the unweighted and investment-weighted means of undistributed TVPI net of DPI, and the extremity of the upper tail for each performance measure: (11) dpi_{p99} and (12) $tvpi_{p99}$ top 1% of DPI; and TVPI; (13) dpi_{tail}, (14) $tvpi_{tail}$ top 1% to the top 10%, (15) $tvpi_{9910}$ top 1% minus bottom 10%. 95% Confidence Intervals shown. Regressions are weighted by the count of entities (portfolio firms) in each sector bin.



(a) Investment Share



(b) Capital Share

Figure A7: Return characteristics and industry tangibility (Compustat)

Notes: This figure reports Eq. (1) coefficients of regressing the Compustat-based industry S&M, R&D, and physical capital intensity measures on the return measures from the figure in the previous page (more details there).

A Matching Patents to NETS

A.1 Construction of the match

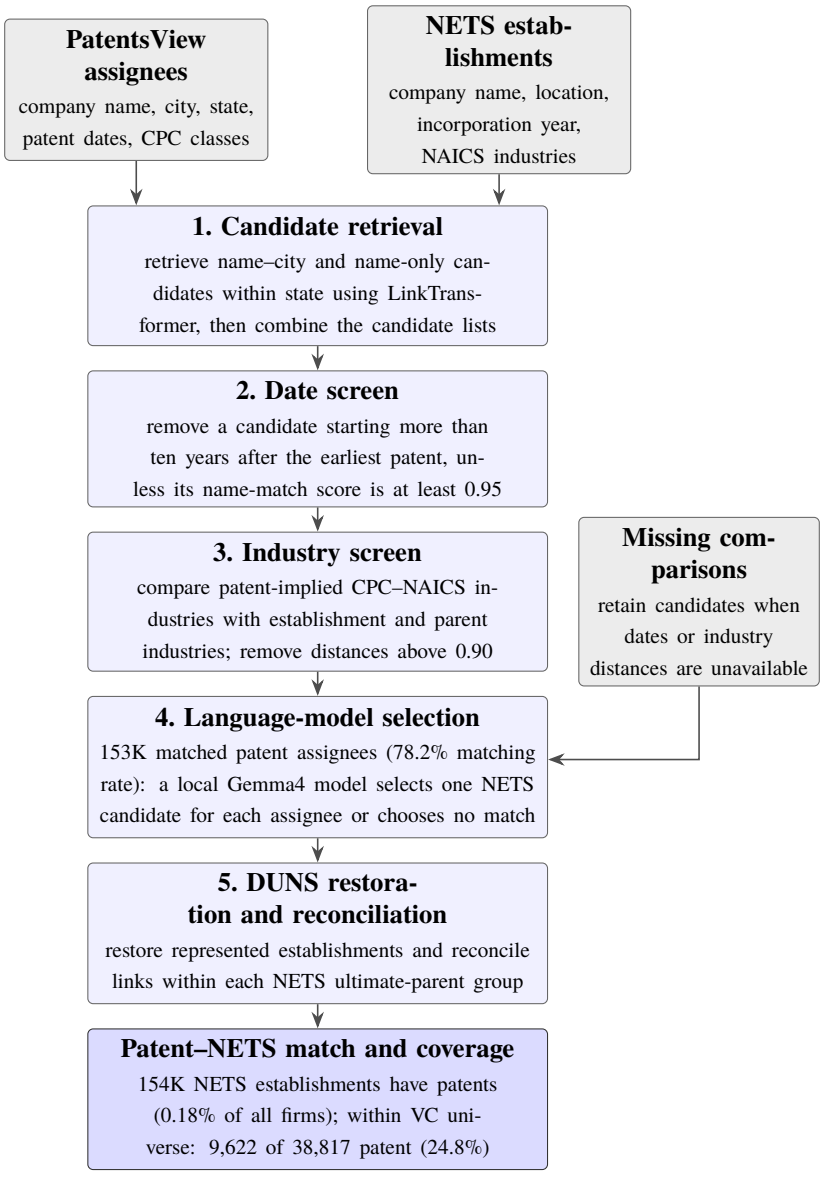
PatentsView assignees are linked to NETS establishments using a four-stage procedure, summarized in Figure A8. First, we use LinkTransformer (Arora and Dell, 2023), a transformer-based record-linkage algorithm that embeds company names and locations in a common vector space and retrieves the most similar NETS records. We use the `dell-research-harvard/lt-wikidata-comp-en` embedding model to retrieve name-city and name-only candidates within the same state, combining the two lists of candidate matches. Second, we compare the assignee's earliest patent date with the establishment's first year in NETS. A patent is dated using the earlier valid filing or grant date. A candidate is removed if the NETS establishment (or its parent company) starts operating more than ten years after the earliest patent, unless the name-match score is at least 0.95. Candidates with missing dates are retained.

Third, we use patent technology classes and NETS industries as an additional screen. Patent CPC subclasses are mapped to four-digit NAICS industries and compared with the industries observed for the candidate establishment (and its parent company) in NETS. Pairs with an industry distance above 0.90 are removed; pairs for which an industry comparison cannot be constructed remain eligible. At this step, we only intend to drop candidate companies that operate in completely different industries – for this purpose we calculate the minimum of pairwise industry distances between Patents data and NETS (measuring the the distance between NETS industries and the closest of recorded patents).

Fourth, a local language model (gemma-4-26B-A4B) is shown the candidates for each assignee and selects one candidate or a no-match option. Company names are used as the main evidence, with city and state used as supporting information.

Selected candidate might be linked to multiple NETS establishments that were collapsed because they had identical name and location. We link the selected candidate back to all underlying DUNS numbers. Within each parent group, we retain the most deeply nested establishment. Establishments linked to different parent groups remain separate matches.

Figure A8: The PatentsView–NETS matching pipeline



A.2 Properties of the match

The LLM evaluates 195,642 assignees and finds a match for 153,053, with a matching rate of 78.2 percent. After applying the NETS hierarchy, the crosswalk contains 160,468 assignee–DUNS links and 154,376 distinct DUNS establishments.

A.3 Longitudinal measures

Each patent associated with an assignee is assigned to every DUNS link retained for that assignee. A patent is dated using the earlier valid filing or grant date, and patents without a valid date are excluded. After it, we compute annual and monthly filing counts and cumulative counts.

B Matching Trademarks to NETS

B.1 Construction of the match

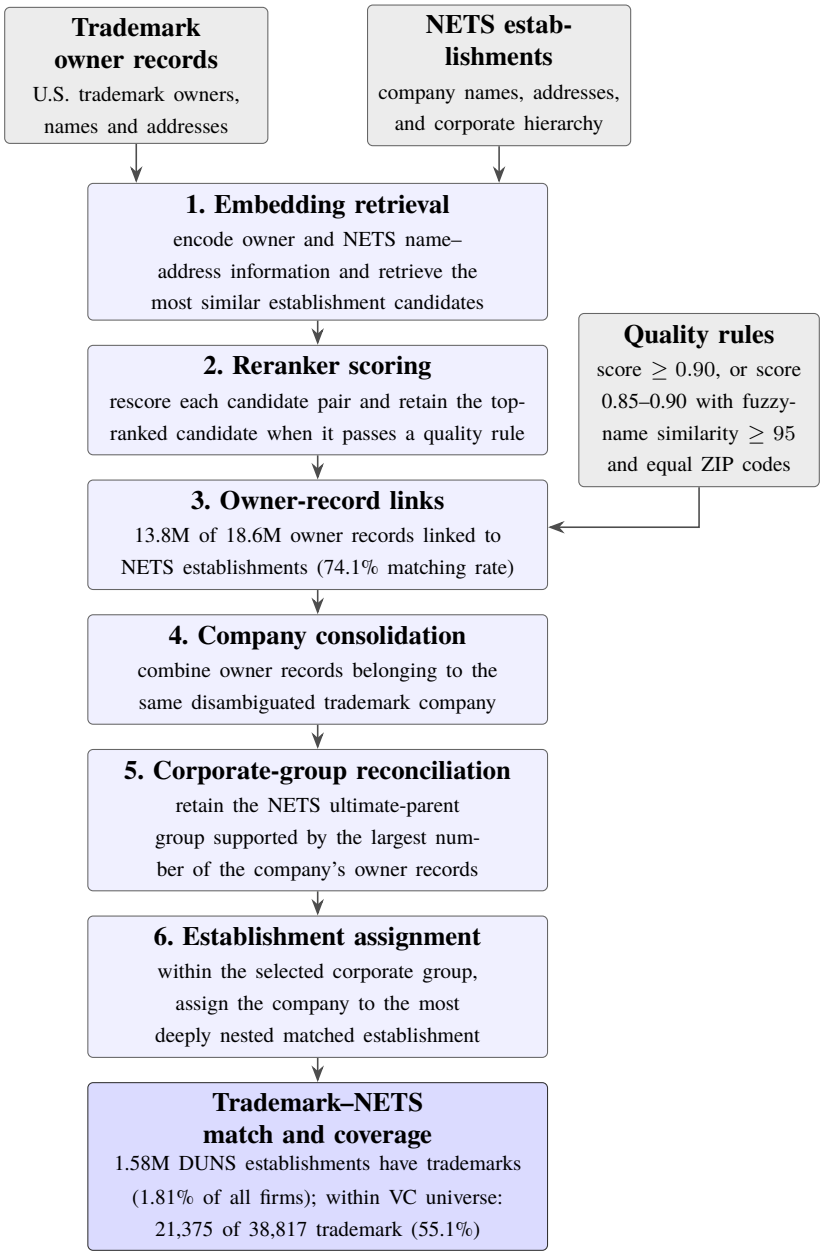
We focus on records with U.S. address whose owner is an organization rather than an individual. The procedure is summarized in Figure A9. In the first matching step, an embedding model uses trademark-owner and NETS name and address information as vectors and retrieves the most similar candidate establishments. A reranker then rescores these candidate pairs. For each owner record, we retain its top-ranked NETS candidate when the pair satisfies one of two quality rules. We keep the candidate if its reranker score is at least 0.9. A candidate with a score between 0.85 and 0.90 is kept only if its fuzzy-name similarity is at least 95 and the ZIP codes are the same.

In the last step, we convert the owner-record links into a company-level match, using two-step procedure. First, owner records are consolidated to disambiguated trademark companies. If the records for one company point to establishments in several NETS corporate groups, we retain the ultimate-parent group supported by the largest number of owner records. Second, within the selected group, we assign the company to the most deeply nested matched establishment.

B.2 Properties of the match

Our crosswalk contains 14,344,490 links. Of the 18,586,000 U.S. organizational-owner records eligible for matching, 13,775,901 are linked, with a matching rate of 74.1 percent. After com-

Figure A9: The USPTO trademark–NETS matching pipeline



pany consolidation and NETS hierarchy reconciliation, the match covers 1,581,768 distinct DUNS establishments.

B.3 Longitudinal measures

Each trademark case is assigned to its trademark filing date when that date is available and otherwise to its earliest subsequent action date. We retain all filings, including marks that are later expired, cancelled, or abandoned. We follow this approach to capture trademark filing activity rather than the stock of trademarks, which allows us to accurately measure the moment when a company decides to start filing trademarks. Then, we calculate filing counts, cumulative counts, and the first observed filing period at the DUNS–year and DUNS–month levels.

C The Revelio Data and the Match to NETS

C.1 The Revelio data

Revelio Labs constructs workforce data from the near-universe of publicly available online professional profiles and resumes, primarily those on LinkedIn. The unit of observation is a position: a worker's employment spell at a company, with a start month, an end month, a job title, a sampling weight, and a predicted salary, together with the Revelio identifier of the employing company. An ongoing spell has a missing end month.

The sampling weight corrects for the fact that coverage of online professional profiles differs across types of workers. Revelio maps each position's job title and location to standard government occupational codes and compares the distribution of observed positions across occupations and geographies with official government employment statistics, drawn from the Bureau of Labor Statistics for the United States. Each position receives a weight such that the reweighted positions match these official employment totals. All firm-level aggregates in this paper are weighted counts and weighted salary sums.

The predicted salary addresses the fact that salaries are not reported on professional profiles. Revelio fits a salary-prediction model on observed salary data from job postings, visa applications, and self-reported sources, and applies the model to every position. The prediction uses position characteristics, such as the job title, seniority level, employer, and location, as well as the worker's

experience.

The cross-sectional analysis uses a point-in-time snapshot: a position is counted as of December 2022 if its start month is on or before December 2022 and its end month is either missing (an ongoing spell) or on or after December 2022.

Revelio also maps each raw job title into a standardized role taxonomy, which is available at several levels of granularity, with rollups of roughly 1,500, 1,000, 500, 150, and 50 detailed roles and, at the top of the hierarchy, a set of broad occupational groups that Revelio labels `k10`. We use this coarsest level, which assigns each position to one of eleven broad groups plus an unknown category for titles that cannot be classified. Table A7 lists the groups and how each is used in this paper: a position is classified as sales and marketing if it is assigned to the "Sales and Marketing" group, and as technical if it is assigned to the "Engineer," "Software Engineer," "Technician," or "Project and IT Specialist" groups. The remaining groups are not assigned to either category but enter the denominator of the wage bill shares.

Table A7: Revelio `k10` role groups and their classification in this paper

Revelio `k10` group	Classification in this paper
Sales and Marketing	Sales and marketing workers
Engineer	Technical workers
Software Engineer	Technical workers
Technician	Technical workers
Project and IT Specialist	Technical workers
Finance	Neither
Operations	Neither
Healthcare Provider	Neither
Public Service and Education	Neither
Service Worker	Neither
Miscellaneous	Neither
Unknown	Neither

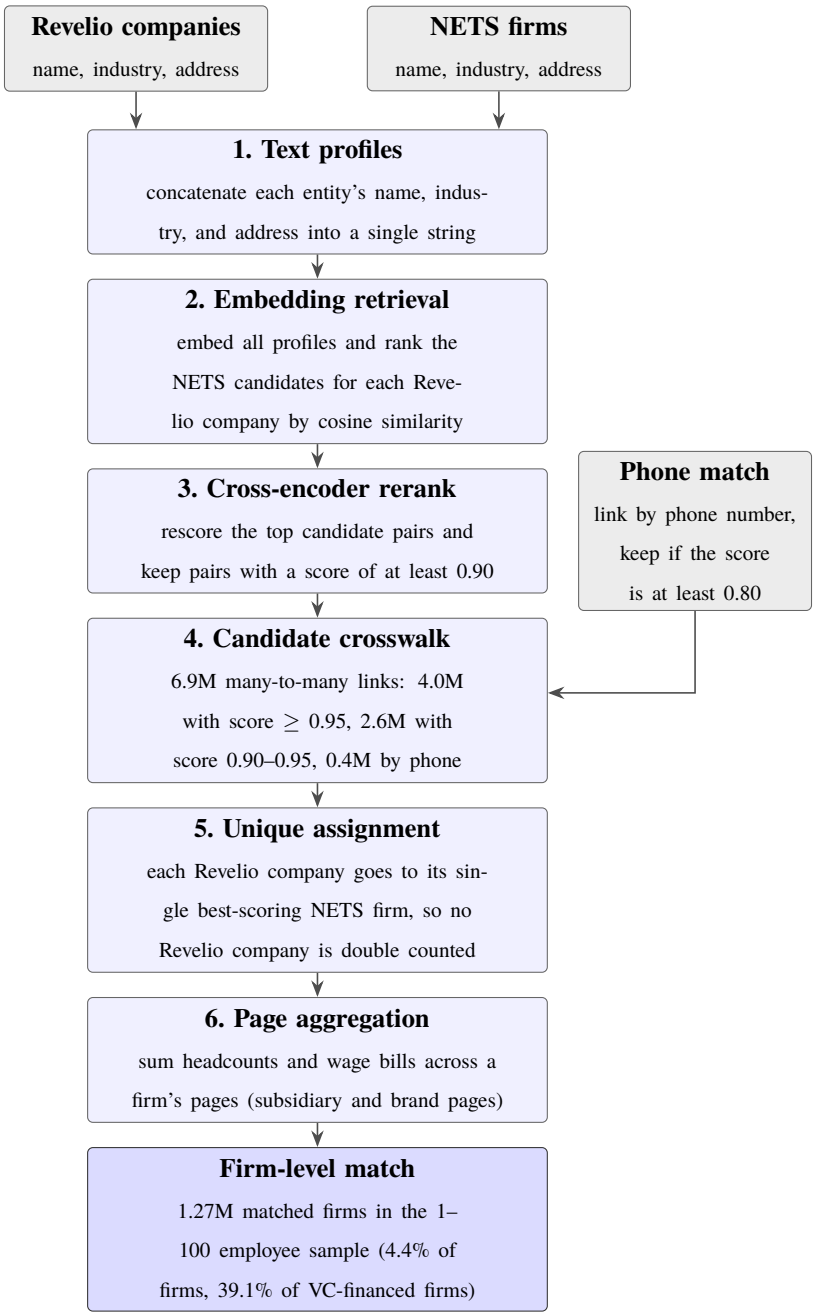
C.2 Construction of the match

Revelio companies are linked to NETS establishments by a multi-stage record-linkage pipeline, summarized in Figure A10. First, a short text profile containing the company name, industry, and address is generated for each entity on both sides. Second, an embedding model encodes all profiles, and for each Revelio company the most similar NETS candidates are retrieved. Third, a cross-encoder model rescores every candidate pair, and pairs with a score of at least 0.90 are kept.[15] Fourth, a supplementary pass matches companies by phone number. Phone matches are kept when the cross-encoder score is at least 0.80. The resulting crosswalk contains 6.9 million candidate links: 4.0 million with a cross-encoder score of at least 0.95, 2.6 million with a score between 0.90 and 0.95, and 0.4 million phone matches.

Two aggregation rules map the resulting links into a firm-level match. A Revelio company that links to several NETS firms is assigned only to the firm with the highest cross-encoder score, so that no Revelio company is double counted across firms. Conversely, one firm can be linked to by several Revelio companies, most commonly subsidiary or brand LinkedIn pages of the same business. The headcounts and wage bills of these pages are summed to the firm level.

[15] The two models play complementary roles. The embedding model encodes each profile independently into a vector, so similarity scores can be computed cheaply over tens of millions of candidate pairs, but the two texts of a pair never interact and the scores are only precise enough for a first screen. The cross-encoder reads a candidate pair jointly and is much more accurate at deciding whether two records describe the same firm, but it is too costly to run on every possible pair. The pipeline therefore uses the embedding model to narrow the candidates and the cross-encoder to make the final decision. The embedding model is `ibm-granite/granite-embedding-small-english-r2` (https://huggingface.co/ibm-granite/granite-embedding-small-english-r2) and the cross-encoder is `Alibaba-NLP/gte-reranker-modernbert-base` (https://huggingface.co/Alibaba-NLP/gte-reranker-modernbert-base).

Figure A10: The NETS–Revelio matching pipeline



Notes: Steps 1–4 build the candidate crosswalk from company names, industries, and addresses, with the phone-number pass as a supplementary source of links. Steps 5–6 turn the many-to-many links into a one-to-one firm-level match. Details and model references are in the text of this appendix section.

C.3 Properties of the match

A firm enters the Revelio-matched sample if the crosswalk links it to a Revelio company with at least one position active in December 2022. Matching therefore requires that at least one of the firm's employees maintains a public professional profile listing an ongoing job at the firm, and that the firm's name and address information allows the crosswalk to identify it in NETS. Of the 28.6 million firms in the 2022 full sample with between 1 and 100 employees, 1.27 million (4.4 percent) are matched. When firms are weighted by employment, the match rate is 12.4 percent. Firm size is by far the strongest predictor of matching: raw match rates rise from about 2 percent for firms with one employee to 39 percent for firms with 51 to 100 employees, and controls attenuate this gradient only modestly, from 36 to 29 percentage points (Table A3). Match rates also rise with firm age, from 1.9 percent for the youngest firms to 11.2 percent for firms aged 30 and older, but most of this raw gradient reflects size: conditional on size, the oldest age band matches only about 3 percentage points more than the youngest.

As discussed in the main text, 39.1 percent of VC-financed firms (19,250 of 49,211) are matched, and VC financing predicts matching far beyond size, age, and industry (Table A3). The VC-financed firms that do not match skew toward one-to-two-employee, very young firms and toward administrative services, whereas the matched ones skew toward manufacturing, information, and professional services. Table A2 compares the full VC-financed sample with its matched subset across the size, age, geography, and deal-timing distributions. The matched VC-financed firms are therefore the LinkedIn-active subset of the VC population. These firms are selected in part on exactly the kind of professional-workforce activity that the outcome variables measure. This is one reason the analysis verifies, wherever the patent and trademark data allow, that results are similar in the full sample that does not require a Revelio match.

D The Non-business Classifier

NETS aims to cover every establishment with a DUNS number, so it includes many records that are not traditional for-profit businesses, such as government units, public libraries, public schools, churches, charities, and member-owned cooperatives. Left in the data, these records would enter the universe as firms that never receive venture financing, inflating the firm counts in the denominators of the venture capital shares and entering the comparison group of the cross-sectional analysis. We

therefore flag them with a text classifier based on the firm's name and drop them from the sample.

The classifier runs on a candidate pool rather than on the full universe. The pool is built from the firm name alone, in one pass over the uppercase NETS name of every firm, and each name faces two tests. The first test checks whether the name contains any of 14 short character strings from an earlier keyword screen. These strings match anywhere in the name, even inside longer words, so LIB catches LIBRARY but also LIBERTY, and the test is deliberately over-inclusive. The second test checks the name against a dictionary of roughly 150 words, abbreviations, and phrases covering government, education, religion, nonprofit, health and social service, and membership organizations. Most dictionary terms count wherever they appear in the name. Short abbreviations such as LIB, TWP, DST, and HOA count only when they stand as a word of their own, and phrases such as CITY OF and FRIENDS OF must appear word for word. A firm enters the pool if its name passes either test, and the keywords play no further role after that.

The screen is designed for recall rather than precision. A false positive is harmless at this stage because every pooled name is re-examined by the classifier, whereas a name that passes neither test is excluded permanently. In a validation on 1,000 randomly sampled firms labeled with the same language-model protocol described below, the screen captured 91 percent of true non-businesses at a precision of only 21 percent. The pool holds 1.5 million candidate firms, about 5 percent of the 29.0 million firms in the master file. Restricting attention to these candidates keeps the classifier calibrated on the kinds of names it was trained on. The resulting flag is then reliable in one direction only: a flagged firm should be dropped, while an unflagged firm may simply be one the screen never selected.

Candidate names are labeled by a large language model, GPT-5.6 Luna. We draw a sample of 10,000 candidates from the pool and split it into batches of 50. Each batch is sent to the model together with a fixed instruction and, for every name in the batch, the firm's name, 3-digit industry code, and employment band. Under a strict output schema, the model returns a six-way label (business, government, public library, education, other nonprofit, or ambiguous), a binary business-status call, a flag for borderline names, and a short reason. The labeled sample is randomly split 80/20 into a training set and a test set, and the training set is used to fine-tune a binary text classifier initialized from ModernBERT, with the firm's name and industry code as inputs. The fitted model assigns each candidate a score, interpreted as the probability that the record is not a business. Candidates scoring at least 0.5 are flagged as non-businesses.

The screen flags 297,819 firms, about 1 percent of the master file. The one-to-two-employee

bands, which contribute 21.4 million of the 29.0 million records in the master file, account for 178,545 of the flagged firms. Together with the removal of publicly listed firms, the screen takes the master file from 29.0 million to 28.7 million firms. The restriction to firms with between 1 and 100 employees then yields the full sample of 28.6 million firms used in the paper.

E Construction of the IPO Sample

E.1 The Ritter sample

The IPO analysis starts from Jay Ritter's IPO database, which lists 14,347 U.S. initial public offerings made between 1975 and 2025. Each offering carries an offer date, the issuer's name and ticker, the firm's founding year, and an independent indicator of venture backing. The population studied here is built from that file in three steps.

The first step is the sample period. We keep offerings made in 2010 or later, of which there are 4,068.

The second step is the link to Compustat. Ritter reports a CRSP permanent number for most offerings, and that number is matched to a Compustat `gvkey` through the CRSP–Compustat Merged linking table, keeping links whose validity window, widened by two months at each end, contains the offer date. This assigns a `gvkey` to 2,764 of the 4,068 offerings, or 68 percent.

The 1,304 that go unassigned fall into two groups. The first group is recent. Ritter fills in CRSP permanent numbers with a lag, and in the vintage of his file used here the field is empty for every offering made after 2023, so none of those 565 offerings can be linked at all. The accounting data would not carry them in any case, since the Compustat extract used here ends in fiscal 2023.

The second group is not recent but is of a particular kind. Of the 739 unmatched offerings made before 2024, 91 percent were offered in 2020 or later, 458 in 2021 alone, and 363, or about half, carry the name of a blank check acquisition vehicle, against 9 percent of the matched offerings. The link therefore fails disproportionately for issuers that are not operating companies, and for whom the intangible intensity measures studied below are not defined.

The third step moves from the offering to the firm. An issuer that lists more than once is counted once, at its earliest offering, and 25 issuers whose first listing predates 2010 are dropped, since for them the offering observed in the sample period is a relisting rather than a new listing. This leaves 2,738 firms, all of which first listed between 2010 and 2023, and it is the population from which

the estimation sample is drawn.

E.2 The linked data sources

With the population fixed, three kinds of information are attached to each firm. Company financials come from Compustat, supplemented with Capital IQ, which reports the selling and marketing expense line that is absent from Compustat (see He, Mostrom and Sufi (2026)). Financing histories, meaning each venture round the firm raised and the date on which it closed, come from PitchBook and Capital IQ. Workforce composition comes from Revelio.

Every link is made on the Compustat `gvkey`, and every link is settled once, before any outcome is measured. These crosswalks are separate from the NETS–Revelio match of Appendix C. The firms studied here are public, so they carry registered identifiers that private firms do not, and the links can therefore be made on exact identifier matches. No fuzzy name matching is used.

Capital IQ companies are linked through the WRDS `gvkey` bridge, which covers all 2,738 firms. Where a `gvkey` maps to several Capital IQ companies, we keep the one that appears most often in the Capital IQ financial panel. PitchBook companies are linked on the Central Index Key (CIK), the identifier the Securities and Exchange Commission assigns to every filer on EDGAR. Where no CIK match exists we use the company web domain, and failing that a normalized exact company name. This links 2,198 firms, or 80 percent, almost all of them on the CIK. Revelio companies are linked either through the `gvkey` that Revelio carries in its own company file or through an exact match on at least two of CIK, CUSIP, ISIN, SEDOL, and EIN, the federal tax identifier the Internal Revenue Service assigns to an employer. This links 2,246 firms, or 82 percent, nine in ten of them through the Revelio-supplied `gvkey`.

Revelio's unit of observation is a company page rather than a corporate group, so a single issuer often appears as several Revelio companies: the parent, its subsidiaries, and its brands. Because the accounting data are consolidated at the parent, a firm's workforce is taken to be the sum of its own Revelio company and of every Revelio company that reports it as its ultimate parent. This affects 41 percent of the linked firms, and rolls 11,355 Revelio companies up into 2,246 firms.

Outcomes and financing histories are built as separate files and joined only at the estimation stage. How a firm's intangible intensity is measured therefore never depends on whether the firm was venture backed.

Three requirements, each described below, reduce these 2,738 firms to the sample used in the

analysis: observed revenue at the last pre-IPO fiscal year (2,184 firms), revenue of at least $25 million (1,317), and a Revelio workforce match (1,041).

E.3 Venture backing

A firm is classified as venture backed if PitchBook or Capital IQ records a venture round that closed strictly before the offer date, or if Ritter codes the offering as venture backed. The two sources agree for 2,175 of the 2,738 firms and disagree for 563.

Every disagreement was adjudicated individually. For each of the 563 firms a large language model with web search was asked whether the firm had raised venture capital before the offering, drawing on its registration statement, contemporaneous funding coverage, and company databases. Two independent passes were run per firm, and a third fresh pass settled the cases where the first two disagreed. Every verdict is stored with a written evidence record and with direct links to the sources it rests on. Of the 563 verdicts, 473 are recorded at high confidence and one at low confidence. Where the passes could not reach a verdict, the source flags are left as they stand.

The adjudication reclassifies 419 firms, 418 of them from venture backed to not venture backed. The reason is that the deal databases record as venture transactions a range of things that are not venture financing of an operating company, among them the sponsor capital raised by blank check shells, sponsored partnerships, carve-outs, and post-bankruptcy relistings, and the low revenue tail of the population is thick with these. The adjudicated variable is the venture indicator used throughout the IPO analysis. All results are qualitatively similar on the sample that ignores adjudication procedure.

E.4 Financial statements and the revenue restriction

Intangible intensity is measured at the last fiscal year that ends before the offer date, so that it describes the firm as it approaches the market rather than after it has raised public equity. Revenue at such a fiscal year is observed for 2,184 of the 2,738 firms. Of the 554 firms for which it is not, 388 have no pre-IPO fiscal year in Compustat at all, seven in ten of them offerings made in 2020 or later, and the other 166 have a fiscal year but do not report revenue. All revenue is converted to 2022 dollars with the GDP deflator before any screen is applied.

The analysis sample is then restricted to firms with revenue of at least $25 million at the last pre-IPO fiscal year. Of the 2,184 firms with reported revenue, 1,317 clear the threshold and 867,

or 39.7 percent, are dropped.

The purpose of the screen is to isolate the sample to firms that can be described as a 'successful' exit by the investors. Of the 867 dropped firms, 412 report no revenue at all and 515 report less than $1 million, and the median dropped firm reports about $53,000. The industry composition of the dropped firms is highly concentrated: 519 of the 867, or 60 percent, are in NAICS 325, chemical manufacturing, and 509 of those are in NAICS 3254, pharmaceutical and medicine manufacturing. These are clinical-stage biotechnology firms that go public to fund trials before they have a product to sell. Fully 64 percent of them report exactly zero revenue at the last pre-IPO fiscal year, and their median revenue is zero. The screen removes 84.9 percent of all firms in NAICS 325 that would otherwise be eligible.

No other industry accounts for more than a small share of the dropped firms. The next largest groups are computer and electronic product manufacturing with 44 firms, firms with no industry assignment with 34, miscellaneous manufacturing including medical devices with 32, credit intermediation with 25, and real estate with 23. Outside chemical manufacturing the dropped firms include a further 78 with zero revenue, among them 20 blank check acquisition vehicles identifiable by name.

The restriction is therefore best understood as removing pre-revenue biotechnology, together with a smaller residue of shell and pass-through vehicles, rather than as removing small operating companies.

It does not fall evenly across the treatment of interest. Among the dropped firms, 59.6 percent are venture backed, against 40.8 percent of the firms retained. Among the dropped chemical manufacturing firms the venture backed share is 84.6 percent. The sample used below is therefore a sample of firms that were already selling something when they listed, and it excludes the part of the venture universe whose customer capital has not yet been built. This is a restriction on external validity rather than a threat to the comparison within the sample: for the excluded firms the ratios the analysis studies are either undefined or mechanically explosive given the extremely low revenue of these firms.

E.5 The estimation sample

The workforce measures require a Revelio match with enough observed employees to form occupation shares. That is available for 1,041 of the 1,317 firms that clear the revenue screen. Regressions

that use both the accounting ratios and the workforce shares are run on this common sample of 1,041 firms, so that the accounting and workforce results describe the same set of firms. Expense ratios and Tobin's q are winsorized at the 5th and 95th percentiles when the financial panel is built, before any IPO subsetting.

F Sample for Dynamic Analysis

Treated firms are NETS firms whose first venture capital financing round falls in 2010–2023, with no pre-2010 VC-class deal and age at most 15 at the round. Because outcomes are measured in Revelio, two sample restrictions are then applied. Defining t_0 for the year of the firm's first VC round: the first restriction (**R1**) requires the firm to be visible in Revelio in the round year itself (at least one position active in December of t_0); the second (**R2**) requires a balanced Revelio panel — the firm visible in every year from t years before to t years after the round, with the last window year no later than 2024, the last reliable Revelio year. The benchmark sample is the R2 panel with $t = 3$: 12,224 (cohorts 2010–2021), out of 82,909 full-sample firms and 27,975 firms passing R1. Table A9 shows how the sample screens affect the sample size, and Table A8 shows the summary statistics for the treatment sample.

The selection criteria requires the firm to be present in Revelio 3 years prior to a VC investment. As a result, the treatment sample contains firms that are older and larger when receiving their initial VC investment. However, the industry and geography mix of the treatment sample is similar to the full sample of firms that receive a VC investment.

Table A8: Treated sample summary statistics (benchmark $[-3, +3]$, equal weight)

	Mean	Median
Birth year	2013	2014
First VC investment year	2017	2017
Total headcount at e = 0	37.6	11.2
GTM share at e = 0 (%)	34.3	31.7
Technical share at e = 0 (%)	31.1	27.6
P(any trademark filing, [-3,-1]) (%)	26.8	0.0
P(any patent filing, [-3,-1]) (%)	10.1	0.0
Firms	12,224	

Notes: Event time e counts years relative to the firm's first clean VC round: $e = 0$ is the round year.

Table A9: Sample restrictions and treated-firm counts

	Treated firms	% of full sample
NETS firm with a PB/CIQ deal (VC or PE class)	190,907	–
has a clean primary venture round	118,487	–
first clean round in 2010–2023	90,180	–
no pre-2010 VC-class deal	90,160	–
Full sample: age at first VC $\leq$ 15	82,909	100.0
R1: in Revelio in year of investment ($t = 0$)	27,975	33.7
R2: balanced Revelio panel $[-3, +3]$	12,224	14.7

Notes: One row per NETS firm. A clean primary venture round is a completed PitchBook deal with DealClass 'Venture Capital' and a primary-round deal type (Capital IQ first rounds, which carry no deal-type detail, pass unscreened). The balanced panel requires presence in Revelio in every year of $[-3, +3]$ around the round with $t_0 + 3 \leq 2024$, so it contains cohorts 2010–2021.